

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 16:25:52 ; Search time 623 Seconds  
(without alignments)  
10334.612 Million cell updates/sec

Title: US-09-923-684-2  
Perfect score: 2859  
Sequence: 1 attcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2859	100.0	2859	24	Human single minde
2	1661.6	58.1	3920	24	Human single minde
3	1625.6	56.9	3885	24	Prostate cancer-as
4	1369	47.9	1464	24	Prostate cancer-as
5	1323.4	46.3	3614	18	Murine SIM (single
6	704.6	24.6	4557	23	DNA encoding novel
7	465.6	16.3	472	24	Unigene #2 used in
8	452	15.8	1905	23	Drosophila melanog
9	231.4	8.1	238	18	Partial human SIM

10	230.6	8.1	711	22	AAK82492	Human immune/haema
11	230.6	8.1	711	22	AAK82495	Human immune/haema
12	230.6	8.1	50000	24	ABK98944	Human asthma-assoc
13	228.8	8.0	11746	22	AAK73167	Human immune/haema
14	228.8	7.9	10612	24	ABK69935	Human secreted pro
15	225.2	7.9	10614	22	AAK36946	Human musculoskele
16	225.2	7.9	10614	22	AAK36946	Human musculoskele
17	224.4	7.8	11337	22	AAK36286	Human polynucleoti
18	224.4	7.8	7818	22	AAK31950	Human immune/haema
19	224	7.8	42519	22	AAK81318	Human reproductive
20	223.8	7.8	1743	22	AAK03368	Human breast or ov
21	223.8	7.8	1743	22	AAI62586	Human reproductive
22	223.8	7.8	1746	22	AAK03369	Human breast or ov
23	223.8	7.8	1746	22	AAI62587	Human pancreatic c
24	223.8	7.8	8676	22	AAK07159	Human cancer relat
25	223.8	7.8	31931	22	ABA07160	Human musculoskele
26	223	7.8	606	24	ABN62170	Genomic sequence #
27	222.8	7.8	8207	22	AAK36305	Human ovarian and
28	222.8	7.8	8207	22	AAK39844	Human reproductive
29	222.8	7.8	8207	22	AAK90215	Human ovarian and
30	222.6	7.8	32134	22	ABA07813	Human reproductive
31	222.6	7.8	32134	22	AAK03615	Human cDNA differe
32	222.6	7.8	32191	22	ABA07814	Human immune/haema
33	222.6	7.8	32191	22	AAK03616	Human immune/haema
34	222.4	7.8	4125	22	AAK84260	Human insulin rece
35	222.4	7.8	4125	22	AAK84261	Human insulin rece
36	222.4	7.8	43599	24	ABK84242	Human insulin rece
37	222.2	7.8	6784	22	AAK69691	Human g35018 cDNA
38	222.2	7.8	7240	22	AAK69693	Human g35018 cDNA
39	222.2	7.8	7240	22	AAH50570	Human g35018 cDNA
40	222.2	7.8	7240	22	AAH31170	Human g35018 cDNA
41	222.2	7.8	7240	22	AAH31267	Human g35018 cDNA
42	222.2	7.8	7240	22	AAH04467	Human g35018 cDNA
43	222.2	7.8	7240	24	ABO72725	Human g35018 cDNA
44	221.8	7.8	1301	21	AAH51627	Human g35018 cDNA
45	221.8	7.8	1386	21	AAH51631	Human g35018 cDNA

## ALIGNMENTS

RESULT 1  
AAD30523  
ID AAD30523 standard; DNA; 2859 BP.

XX AAD30523;

AC AAD30523;

DT 31-MAY-2002 (first entry)

DE Human single minded homologue 2 (SIM2) gene, short form.

DE Human: single minded homologue 2; SIM2; colon cancer; prostate cancer;

KW pancreas cancer; chromosome 21q22.2; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 93..1805

FT /\*tag= a

FT /product= "Human SIM2 protein, short form"

XX WO200212565-A1.

XX 14-FEB-2002.

XX 06-AUG-2001; 2001WO-US24781.

XX 04-AUG-2000; 2000US-223531P.

XX 22-DEC-2000; 2000US-257965P.

XX (UYFL ) UNIV FLORIDA ATLANTIC.

XX Narayanan R;

XX WPI; 2002-227169/28.  
DR P-PSDB; AAE19388.  
XX  
PT Detecting cancer in tissue sample, involves providing tissue sample,  
PT and analyzing tissue sample for presence of single minded homolog 2  
PT marker  
XX  
XX Claim 9; Page 55-56; 60pp; English.  
XX  
XX The invention relates to a method of detecting cancer in a tissue  
CC sample. The method involves providing the tissue sample and analyzing  
CC the tissue sample for the presence of a single minded homolog 2 (SIM2)  
CC nucleic acid or protein, where the presence of the SIM2 marker in the  
CC tissue sample indicates that the tissue sample contains cancer. The  
CC tissue sample is useful for identifying compounds that modulate expression of  
CC SIM2 gene in a cell. It is useful for detecting and reducing the growth  
CC of cancer such as colon, prostate and pancreas cancer in an animal or  
CC mammal. The present sequence is human SIM2 gene, short form. SIM2 gene  
CC is located on chromosome 21q22.2.  
XX  
XX Sequence 2859 BP; 701 A; 803 C; 778 G; 577 T; 0 other;  
XX  
XX Query Match 100.0%; Score 2859; DB 24; Length 2859;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 2859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACTCACTATAGGGCTCGAGCGCCGCCGGGAGGTGGGGCTCCCGGGCTGGAGCAGC 60  
DB 1 ACTCACTATAGGGCTCGAGCGCGCCGCCGGGAGGTGGGGCTCCCGGGCTGGAGCAGC 60  
QY 61 GCCGGGTCTAATATGCGCGGAGCGGAGCGCGGATGAAGGAGATCCAAAGATGGGCCA 120  
DB 61 GCCGGGTCTAATATGCGCGGAGCGGAGCGCGGATGAAGGAGATCCAAAGATGGGCCA 120  
QY 121 AGACAGGAGGAGAGAGAAATGGCGAGTTTACAGGCTTGCCAGGCTGCCAGCTGCCGCTGC 180  
DB 121 AGACAGGAGGAGAGAAATGGCGAGTTTACAGGCTTGCCAGGCTGCCAGCTGCCGCTGC 180  
QY 181 CGTCGGCCATCATTGCGAGCTGGACAAAGCGTCCATATCGCTCCACAGAGTACC 240  
DB 181 CGTCGGCCATCATTGCGAGCTGGACAAAGCGTCCATATCGCTCCACAGAGTACC 240  
QY 241 TGAAGATGCGCGCTCTCCCGAAGTTTAGAGACGCGTGGGGAGCAGCGAGCGCGG 300  
DB 241 TGAAGATGCGCGCTCTCCCGAAGTTTAGAGACGCGTGGGGAGCAGCGAGCGCGG 300  
QY 301 CCGGGCCCTGGAGCGGCTGCCAAGAGCTGGGATCGGATCGGATCGGATCGGATCGGATG 360  
DB 301 CCGGGCCCTGGAGCGGCTGCCAAGAGCTGGGATCGGATCGGATCGGATCGGATCGGATG 360  
QY 361 GATTTGTTTTTGTGTAGCATCTGATGGCAAAATCATATGATATATCCGAGACCGCTTCTG 420  
DB 361 GATTTGTTTTTGTGTAGCATCTGATGGCAAAATCATATGATATATCCGAGACCGCTTCTG 420  
QY 421 TCCATTTAGGCTTATCCAGGTGAGCTCAGCGGCAACAGTATTTATGATATCATCATC 480  
DB 421 TCCATTTAGGCTTATCCAGGTGAGCTCAGCGGCAACAGTATTTATGATATCATCATC 480  
QY 481 CTTCTGACACGATGAGATGACCGCTGCTCAGCGGCCACAGCGCGCTGCACACACAC 540  
DB 481 CTTCTGACACGATGAGATGACCGCTGCTCAGCGGCCACAGCGCGCTGCACACACAC 540  
QY 541 TGCCTCCAAGAGTATGAGATGAGAGGTGCTTTCTTCGTAATGATGCTGTGGCGA 600  
DB 541 TGCCTCCAAGAGTATGAGATGAGAGGTGCTTTCTTCTTCTGAATGATGCTGTGGCGA 600  
QY 601 AAAGAACCGGGCTGACCTGCAGCGGATACAGGTCATCATCTGAGTGGCTTACTTGA 660  
DB 601 AAAGAACCGGGCTGACCTGCAGCGGATACAGGTCATCATCTGAGTGGCTTACTTGA 660  
QY 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTACGACTCTCTGTACCAGATTGTGGGC 720  
DB 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTACGACTCTCTGTACCAGATTGTGGGC 720

DB 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTACGACTCTCTGTACCAGATTGTGGGC 720  
QY 721 TGGTGGCGTGGCCAGTTCGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780  
DB 721 TGGTGGCGTGGCCAGTTCGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780  
QY 781 ACATCTTCATGTTTCAGGGCCAGCCTTGACCTGAAGCTGATATTCCTGGATTCAGGGTGA 840  
DB 781 ACATCTTCATGTTTCAGGGCCAGCCTTGACCTGAAGCTGATATTCCTGGATTCAGGGTGA 840  
QY 841 CCGAGGTGACGGGTTACGAGCCGAGGACCTGTGAGAGACCTTATACCATCACGTGC 900  
DB 841 CCGAGGTGACGGGTTACGAGCCGAGGACCTGTGAGAGACCTTATACCATCACGTGC 900  
QY 901 ACGGCTGCCAGCTGTTTCCACCTCCGCTACGACACACCTCCTCTGTGTGAAGGCCAGG 960  
DB 901 ACGGCTGCCAGCTGTTTCCACCTCCGCTACGACACACCTCCTCTGTGTGAAGGCCAGG 960  
QY 961 TCACCAACCAAGTACTACCGGCTGCTTCCACGCGGGGCTGGGTGGGTGTCAGAGCT 1020  
DB 961 TCACCAACCAAGTACTACCGGCTGCTTCCACGCGGGGCTGGGTGGGTGTCAGAGCT 1020  
QY 1021 ACGCCACCGTGTGCAACAACAGCCGCTGCTCCCGGCCCTACCTGCTGATGCTCAATT 1080  
DB 1021 ACGCCACCGTGTGCAACAACAGCCGCTGCTCCCGGCCCTACCTGCTGATGCTCAATT 1080  
QY 1081 ATGACTCAGGAGATGATACAGGAACTTACAGTGTCCCTGGAGCAGGTGTCCACTG 1140  
DB 1081 ATGACTCAGGAGATGATACAGGAACTTACAGTGTCCCTGGAGCAGGTGTCCACTG 1140  
QY 1141 CCAAGTCCCAGGACTCCTGGAGGACCGCTTGTCTACCTCAACAAGAACTAGAAATTAG 1200  
DB 1141 CCAAGTCCCAGGACTCCTGGAGGACCGCTTGTCTACCTCAACAAGAACTAGAAATTAG 1200  
QY 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAACTTACCCCCACAGC 1260  
DB 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAACTTACCCCCACAGC 1260  
QY 1261 AATACAGCTCATTTCCAAATGGACAACTGGAATGCGGCCAGCTCGGAACTGGAGAGCCA 1320  
DB 1261 AATACAGCTCATTTCCAAATGGACAACTGGAATGCGGCCAGCTCGGAACTGGAGAGCCA 1320  
QY 1321 GTCCCGCTTGAAGCGCTGCTCCTCCAGAACTGCAAGCCCACTCAGAAAGCAGTGACC 1380  
DB 1321 GTCCCGCTTGAAGCGCTGCTCCTCCAGAACTGCAAGCCCACTCAGAAAGCAGTGACC 1380  
QY 1381 TTCTGTACAGCCATCTTACAGCTGCGCTTCTCTACCATTTAGACACTTCCTCTGG 1440  
DB 1381 TTCTGTACAGCCATCTTACAGCTGCGCTTCTCTACCATTTAGACACTTCCTCTGG 1440  
QY 1441 ACTCTCACTTCTTACAGCAGCAAAAGCAATGTTGCGGCCCAAGTTCGGCAGCCCAAG 1500  
DB 1441 ACTCTCACTTCTTACAGCAGCAAAAGCAATGTTGCGGCCCAAGTTCGGCAGCCCAAG 1500  
QY 1501 GATCCCTTGTGAGCTGGCAGCTTTTCTTCCAGCACAATGCCAGCCAGCGGTGAATGCC 1560  
DB 1501 GATCCCTTGTGAGCTGGCAGCTTTTCTTCCAGCACAATGCCAGCCAGCGGTGAATGCC 1560  
QY 1561 AGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCCTCTCCAGCTAAAATCCTCCAG 1620  
DB 1561 AGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCCTCTCCAGCTAAAATCCTCCAG 1620  
QY 1621 AGCCACCGGCAACACTGCTAGGCAGCTGCTGCGCAAGCTTACGAAGGTGGGTGAGTGC 1680  
DB 1621 AGCCACCGGCAACACTGCTAGGCAGCTGCTGCGCAAGCTTACGAAGGTGGGTGAGTGC 1680  
QY 1681 TGTCTGTGGGAGGTGGGAGGACTGCGCACCGCGGGAGCGGAGAGAGCCATGCGGTG 1740  
DB 1681 TGTCTGTGGGAGGTGGGAGGACTGCGCACCGCGGGAGCGGAGAGAGCCATGCGGTG 1740  
QY 1741 GGTGGCAGATGGAGACAGAACCTTACGCTTGGGCAAACTTGGCCCTCTTTCTGTTCTA 1800  
DB 1741 GGTGGCAGATGGAGACAGAACCTTACGCTTGGGCAAACTTGGCCCTCTTTCTGTTCTA 1800



181	CGTGGCCATCCTTCAGCTGACAAAGCGTCCATCATCCGCTCACCAGAGTACC	240	
182			
183			
184			
185			
186			
187			
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200			
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211			
212			
213			
214			
215			
216			
217			
218			
219			
220			
221			
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247			
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			
262			
263			
264			
265			
266			
267			
268			
269			
270			
271			
272			
273			
274			
275			
276			
277			
278			
279			
280			
281			
282			
283			
284			
285			
286			
287			
288			
289			
290			
291			
292			
293			
294			
295			
296			
297			
298			
299			
300			
301			
302			
303			
304			
305			
306			
307			
308			
309			
310			
311			
312			
313			
314			
315			
316			
317			
318			
319			
320			
321			
322			
323			
324			
325			
326			
327			
328			
329			
330			
331			
332			
333			
334			
335			
336			
337			
338			
339			
340			
341			
342			
343			
344			
345			
346			
347			
348			
349			
350			
351			
352			
353			
354			
355			
356			
357			
358			
359			
360			
361			
362			
363			
364			
365			
366			
367			
368			
369			
370			
371			
372			
373			
374			
375			
376			
377			
378			
379			
380			
381			
382			
383			
384			
385			
386			
387			
388			
389			
390			
391			
392			
393			
394			
395			
396			
397			
398			
399			
400			
401			
402			
403			
404			
405			
406			
407			
408			
409			
410			
411			
412			
413			
414			
415			
416			
417			
418			
419			
420			
421			
422			
423			
424			
425			
426			
427			
428			
429			
430			
431			
432			
433			
434			
435			
436			
437			
438			
439			
440			
441			
442			
443			
444			
445			
446			
447			
448			
449			
450			
451			
452			
453			
454			
455			
456			
457			
458			
459			
460			
461			
462			
463			
464			
465			
466			
467			
468			
469			
470			
471			
472			
473			
474			
475			
476			
477			
478			
479			
480			
481			
482			
483			
484			
485			
486			
487			
488			
489			
490			
491			
492			
493			
494			
495			
496			
497			
498			
499			
500			
501			
502			
503			
504			
505			
506			
507			
508			
509			
510			
511			
512			
513			
514			
515			
516			
517			
518			
519			
520			
521			
522			
523			
524			
525			
526			
527			
528			
529			
530			
531			
532			
533			
534			
535			
536			
537			
538			



CC The present invention relates to methods of detecting a prostate  
 CC cancer-associated transcript in a cell from a patient. The method  
 CC comprises contacting a biological sample from the patient with  
 CC prostate cancer-associated polynucleotides (designated PC genes) that  
 CC selectively hybridize to a sequence that is at least 80% identical  
 CC to them. The prostate cancer-associated polynucleotide sequences  
 CC are differentially expressed in prostate tumour tissue or in  
 CC prostate cancer and are derived from the tissues of various  
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
 CC The methods of the invention are useful for diagnosing and treating  
 CC prostate cancer in mammals. The prostate cancer-associated genes are  
 CC useful for diagnosing or treating prostate cancer, as well as for  
 CC identifying modulators of prostate cancer or agents that inhibit  
 CC prostate cancer. The nucleic acid sequences are particularly useful  
 CC in gene therapy, as a vaccine or in antisense applications.  
 CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide  
 CC sequences.  
 XX

Sequence 3885 BP; 916 A; 1125 C; 985 G; 859 T; 0 other;

Query Match		56.9%;	Score 1625.6;	DB 24;	Length 3885;
Best Local Similarity		99.8%;	Pred. No. 0;		
Matches 1628;		Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	37	GGGGCTCCGGCGCTGGAGCAGCGCGGGTCTAATATGCCGGAGCGCGCGATGA	96		
DB	1	GGGGCTCCGGCGCTGGAGCAGCGCGGGTCTAATATGCCGGAGCGCGCGATGA	60		
QY	97	AGGAGAAGTCCAAAGATCGGCAAGACGAGGAGGAGGAGAAATGGCGAGTTTACG	156		
DB	61	AGGAGAAGTCCAAAGATCGGCAAGACGAGGAGGAGGAGAAATGGCGAGTTTACG	120		
QY	157	AGCTTGCCTGCTCCGGTGGCGGCTGCGGCATCATCTTCAGCTGCGACAAAGCGTCCA	216		
DB	121	AGCTTGCCTGCTCCGGTGGCGGCTGCGGCATCATCTTCAGCTGCGACAAAGCGTCCA	180		
QY	217	TCATCCGCTCACCACAGCTACCTGAAGATGCGCGCGCTTCCCGAAGGTTTAGGAG	276		
DB	181	TCATCCGCTCACCACAGCTACCTGAAGATGCGCGCGCTTCCCGAAGGTTTAGGAG	240		
QY	277	ACGGCTGGGACAGCG	336		
DB	241	ACGGCTGGGACAGCG	300		
QY	337	CGCACTTCTGCAGACTTTGGATGATTTGTTTGGTAGCATCTGATGGCAAAATCA	396		
DB	301	CGCACTTCTGCAGACTTTGGATGATTTGTTTGGTAGCATCTGATGGCAAAATCA	360		
QY	397	TGTATATATCCGAGACCGCTTCTGTCCATTTAGCTTTATCCAGGTGGAGCTCACGGCA	456		
DB	361	TGTATATATCCGAGACCGCTTCTGTCCATTTAGCTTTATCCAGGTGGAGCTCACGGCA	420		
QY	457	ACAGTATTTATGAATACATCCCTTCTGACACAGATGAGATGACCGTGTCTCACGG	516		
DB	421	ACAGTATTTATGAATACATCCCTTCTGACACAGATGAGATGACCGTGTCTCACGG	480		
QY	517	CCACACGCGCTGCACACACCTGCTCCAAAGATATGAGATGAGAGGTGCTTCTTC	576		
DB	481	CCACACGCGCTGCACACACCTGCTCCAAAGATATGAGATGAGAGGTGCTTCTTC	540		
QY	577	TTGCAATGAATGTCTTGGCGAAAGAAAGCGCGCGCTGACCTGCAGCGGATACAAGG	636		
DB	541	TTGCAATGAATGTCTTGGCGAAAGAAAGCGCGCGCTGACCTGCAGCGGATACAAGG	600		
QY	637	TCATCCACTGCAGTGGCTACTTGAAGATCAGCAGTATATGCTGGACATGCTCCGTACG	696		
DB	601	TCATCCACTGCAGTGGCTACTTGAAGATCAGCAGTATATGCTGGACATGCTCCGTACG	660		
QY	697	ACTCTCTACAGATTTGGGGCTGGTGGCGCTGGCGCAGTGCCTGCGCACCGCTGCGCA	756		
DB	661	ACTCTCTACAGATTTGGGGCTGGTGGCGCTGGCGCAGTGCCTGCGCACCGCTGCGCA	720		
QY	757	TCACCGAGATCAAGCTGTACATACATGTTTCATGTTTCAGGGCCAGCGCTTGACCTGAAGC	816		

DB	721	TCACCGAGATCAAGCTGTACAGTAACTGTTTCATGTTTCAGGGCCAGCGCTTGACCTGAAGC	780
QY	817	TGATATTTCTCGATTCCAGGGTGACCGAGGTGACGGGTACGAGCCCGCAGGACCTGATCG	876
DB	781	TGATATTTCTCGATTCCAGGGTGACCGAGGTGACGGGTACGAGCCCGCAGGACCTGATCG	840
QY	877	AGAAGACCTATACCATCATCGTGCACGGCTGCGACGCTGTTTCCACCTCCCGTACGACAC	936
DB	841	AGAAGACCTATACCATCATCGTGCACGGCTGCGACGCTGTTTCCACCTCCCGTACGACAC	900
QY	937	ACCTCTCTGTTGGTGAAGGGCCAGGTACACCAAGTACTACCGCTGCTGTCTCAAGCGGG	996
DB	901	ACCTCTCTGTTGGTGAAGGGCCAGGTACACCAAGTACTACCGCTGCTGTCTCAAGCGGG	960
QY	997	CGCGCTGGGTGGTGGTGCAGAGTACGCCACCTGGTGCACACACCGCTGCTCCCGGC	1056
DB	961	CGCGCTGGGTGGTGGTGCAGAGTACGCCACCTGGTGCACACACCGCTGCTCCCGGC	1020
QY	1057	CCCACTGCATCGTGAAGTCAATTAATGTTACTACCGAGATTCATTAACAGGAACCTTACG	1116
DB	1021	CCCACTGCATCGTGAAGTCAATTAATGTTACTACCGAGATTCATTAACAGGAACCTTACG	1080
QY	1117	TGTCCTCTGGAGCAGGTGTCCACTGCGCAAGTCCCAAGTCCCGAGGACCGCTTGTCTTA	1176
DB	1081	TGTCCTCTGGAGCAGGTGTCCACTGCGCAAGTCCCGAGGACCGCTTGTCTTA	1140
QY	1177	CCTCAACAAGAACTAGGAATTTAGTGAACACCCAAAATACCAAGATGAAGACAAAGCTGA	1236
DB	1141	CCTCAACAAGAACTAGGAATTTAGTGAACACCCAAAATACCAAGATGAAGACAAAGCTGA	1200
QY	1237	GAACAACCTTACCCCCACACAGCAATACAGCTCATTTCCAATGGACAACTGGAATGCG	1296
DB	1201	GAACAACCTTACCCCCACACAGCAATACAGCTCATTTCCAATGGACAACTGGAATGCG	1260
QY	1297	GCAGCTCGGAACCTGGAGAGCGAGTCCCGCTGCAAGCGCTGCTGCTCCAGAACTGCG	1356
DB	1261	GCAGCTCGGAACCTGGAGAGCGAGTCCCGCTGCAAGCGCTGCTGCTCCAGAACTGCG	1320
QY	1357	AGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCTTACAGCTGCGCTTCTCCT	1416
DB	1321	AGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCTTACAGCTGCGCTTCTCCT	1380
QY	1417	ACCATTTAGGACACTTCCCTCTGACACTCTCACTTCTTACAGCAGCAAAAGCCAAATGTC	1476
DB	1381	ACCATTTAGGACACTTCCCTCTGACACTCTCACTTCTTACAGCAGCAAAAGCCAAATGTC	1440
QY	1477	CGGCAAGTTCCGGCAGCGCCCAAGGATCCCTTGTAGGTGGCAGCGCTTTTCTCTGAGCA	1536
DB	1441	CGGCAAGTTCCGGCAGCGCCCAAGGATCCCTTGTAGGTGGCAGCGCTTTTCTCTGAGCA	1500
QY	1537	CAATGCGACCGCAGCGGTGAATGCGCAGTGGCATATGCAACCCCTAGTGCCTAGCAGCT	1596
DB	1501	CAGTGCAGCGCAGCGGTGAATGCGCAGTGGCATATGCAACCCCTAGTGCCTAGCAGCT	1560
QY	1597	CGTCTCCAGCTTAAATATCTCCAGAGCCACCGGGAACACTGCTAGGACAGCGCTGGTGC	1656
DB	1561	CGTCTCCAGCTTAAATATCTCCAGAGCCACCGGGAACACTGCTAGGACAGCGCTGGTGC	1620
QY	1657	CAAGCTACGAG 1668	
DB	1621	CAAGCTACGAG 1632	

RESULT 4  
 ABK92250  
 ID ABK92250 standard; DNA; 1464 BP.  
 XX  
 AC ABK92250;  
 XX  
 DT 15-AUG-2002 (first entry)  
 XX  
 DE Prostate cancer-associated DNA sequence #136.



```

Db      1260 GCCAGCCAGCGGTGAATGCCAGTGGCATATATGCCAACCCCTAGTGCCTAGCAGCTCGTC 1319
Qy      1601 TCCAGCTAAATTCCTCCAGAGCCACCGCGAACAACACTGCTAGGCACAGCCTGGTGCCAAAG 1660
Db      1320 TCCAGCTAAATTCCTCCAGAGCCACCGCGAACAACACTGCTAGGCACAGCCTGGTGCCAAAG 1379
Qy      1661 CTACGAAGG 1669
Db      1380 CTACGAAGG 1388

RESULT 5
AAAT91883
ID      AAAT91883 standard; cDNA to mRNA; 3614 BP.
XX      AC
XX      AAAT91883;
XX      30-JAN-1998 (first entry)
XX      Murine SIM (single-minded) gene product coding sequence.
DE      Mouse; SIM; single-minded; early development; central nervous system;
KW      CNS; diagnosis; Down's syndrome; N-terminal; ds.
XX      Mus sp.
XX      Key
XX      CDS
XX      Location/Qualifiers
XX      504..2477
XX      /*tag= a
XX      JP09268199-A.
XX      14-OCT-1997.
XX      28-MAR-1996; 96JP-0074261..
XX      29-JAN-1996; 96JP-0012952.
XX      31-MAR-1995; 95JP-0076362.
XX      (FUJI ) FUJISAWA PHARM CO LTD.
XX      WPI; 1997-554711/51.
XX      DR      P-PSDB; AAW30758.
XX      PT      Human or mouse single-minded gene - used in the diagnosis of Down's
XX      syndrome
XX      Claim 5; Pages 12-13; 17pp; Japanese.
XX      This is the murine SIM (single-minded) gene. The SIM gene appears to be
XX      of importance in controlling the early development of the central nervous
XX      system. In particular the gene, and its products are used in the
XX      diagnosis of Down's syndrome.
XX      Sequence 3614 BP; 833 A; 1110 C; 952 G; 717 T; 2 other;
XX      Query Match 46.3%; Score 1323.4; DB 18; Length 3614;
XX      Best Local Similarity 88.2%; Pred. No. 0;
XX      Matches 1451; Conservative 0; Mismatches 191; Indels 3; Gaps 1;
Qy      27 CCGGGCAGGTGGGGCTCCCGGGCTGGAGCAGCGCGGTCTAATATGCCCGGAGCCGA 86
Db      438 CAGCCCAAGCGGGGACTCCCGGGCTGGAGCAGCGAGGTCTAATATGCCCGGAGCCGA 497
Qy      87 GCGCGCATCAAGGAGAGTCCCAAGATCGGGCCAAAGACAGAGGGAGAGGAAATGGC 146
Db      498 GCGCGCATCAAGGAGAGTCCCAAGATCGGGCCAAAGACAGAGGGAGAGGAAATGGC 557.
Qy      147 GAGTTTTAGAGCTTGCCCAAGCTGCTCCCGGTGGCGGTCCGCATCATCTCCAGCTGGAC 206
Db      558 GAGTTTTAGAGCTTGCCCAAGCTGCTCCCGGTGGCGGTCCGCATCATCTCCAGCTGGAC 617

```

```

Qy      207 AAAGCGTCCATCATCCCGCTCACCAGAGCTACTGAGATGCGGCCCGTCTTCCCGGAA 266
Db      618 AAAGCGTCCATCATCCCGCTCACCAGAGCTACTGAGATGCGGCCCGTCTTCCCGGAA 677
Qy      267 GGTTTAGAGACGCGTGGGACAGCGCGCGCGCGCGCTGACCGCGCTGCGCCAG 326
Db      678 GGTCTAGAGACGCGTGGGACAGCGCGCGCGCGCTGACCGCGCTGCGCCAG 737
Qy      327 GAGCTGGGATCGCACTTGTGTCAGACTTTGGATGGAATTTGTTTGTGTAGCATCTGAT 386
Db      738 GAGCTGGGATCGCACTTGTGTCAGACTTAGATGGAATTTGTTTGTGTGCTGCTGAT 797
Qy      387 GGCATAATCATGTATATCCGAGACCGCTCTGTCCATTTAGGCTTATCCAGGTGAG 446
Db      798 GGCATAATCATGTATATCTGAGACCGCTCTGTCCATTTAGGCTTGTCCAGGTGAG 857
Qy      447 CTCACGGGCAACAGTATTTATGAATACATCCATCTTGTGACACAGATGAGATGCCGT 506
Db      858 CTCACAGGCAACAGTATTTATGAATACATACATCTTGTGACACAGATGAGATGCC 917
Qy      507 GTCTCAGCGCCACCGCGCTGACACCGACCTGCTCCCAAGAGTATGAGATGAGAGG 566
Db      918 GTACTTACAGCGCCACCGCGCTCCACCATCACTGCTCCCAAGAGTATGAGATGAGAGG 977
Qy      567 TCGTTCTTTCTTCCGAATGAATGTGTCTTGGCGAAGGAAAGCGCGGCTGACCTGCGAGC 626
Db      978 TCTTTCTTTCTTCCGAATGAATGTGTCTTGGCGAAGGAAAGCGCGGCTGACATGCGAGT 1037
Qy      627 GGATACAGGTCACTCAGCTGAGTGGCTACTTGAAGATCAGCGAGTATATCTGACATG 686
Db      1038 GGATACAGGTCACTCAGCTGAGCGGCTACCTAAAGATCAGACATACATCTGAGCATG 1097
Qy      687 TCCCTGTAGGACTCCTGCTTACCAGATGTGGGGCTGTGGCGGTGGCGGCTGAGCGGTGAGCGGCCGAG 866
Db      1098 TCCCTGTAGGACTCCTGCTTACCAGATGTGGGGCTGTGGCGGTGGCGGCTGAGCGGTGAGCGGCCA 1157
Qy      747 CCCAGTGCATCAGCGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTTCAGGCGCAGCCTT 806
Db      1158 CCCAGCGCATCAGCGAGATCAAGCTCCACAGCAACATGTTTCATGTTTTCAGGCGCAGCCTT 1217
Qy      807 GACCTGAAGCTGATATCTCGATTCCAGGCTGACCGGCTGAGCGGTGAGCGGTGAGCGGCCGAG 866
Db      1218 GACCTGAAGCTGATATCTCGATTCCAGGCTGACCGGCTGAGCGGTGAGCGGCCA 1277
Qy      867 GACCTGATCGAAGAGACCTATACCATCAGCTGACGCGCTCGGACGCTGTTTCCACCTCCGC 936
Db      1278 GACCTGATCGAAGAGACCTATACCATCAGCTGACGCGCTCGGACGCTGTTTCCACCTCCGC 1337
Qy      927 TACGCACACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCAGCTGCTG 986
Db      1338 TATGCACACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCAGCTGCTG 1397
Qy      987 TCCAAGCGGGCGGCTGGGTGTGGGTGCAGAGCTAGCCACCGCTGTCACACACACCGCCG 1046
Db      1398 TCCAAGCTGGCGGCTGGGTGTGGGTGCAGAGCTAGCCACCGCTGTCACACACCGCCG 1457
Qy      1047 TCGTCCCGCGCCACCTCATCTGAGTGTCAATTTATGTTACTACAGGAGTGAATGAACAAG 1106
Db      1458 TCTCCCGGCTCAGTGCATCTGAGTGTCAATTTATGTTCTCACCGATGTTGAATACAAG 1517
Qy      1107 GAACTTCAGCTGTCCCTGGAGCAGGTGTCCACTGCCAAGTCCCGAGGCTCTCTGGAGGACC 1166
Db      1518 GAACTTCAGCTGTCTCTGGAGCAGGTATCCACCTCTAAGTCCCGAGGCTCTCTGGAGGACC 1577
Qy      1167 GCCTTGTCTACCTCACAAGAACTAGAAATTAGTGAACCCCAAAATACCAAGATGAAG 1226
Db      1578 ACCTTGTCTACCTCACAAGAACTAGAAATTAGTGAACCCCAAAATACCAAGATGAAG 1637
Qy      1227 ACAAGCTGAGAAACAACTTACCCCGCCACAGCATACAGCTCATTTCAATGACAAA 1286
Db      1638 ACAAGCTGAGAAACAACTTACCCCGCCACAGCATACAGCTCATTTCAATGACAAA 1697
Qy      1287 CTGGAATGGCGGAGCTCGGAAACTGGAGAGCGGCTGCCCGCTGCAAGCGGCTGCTCCT 1346

```

(II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, for example, gene mapping, identification of mutations in CC responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human CC diagnostic coding sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4557 BP; 1276 A; 1147 C; 1068 G; 1066 T; 0 other;

Query Match 24.6%; Score 704.6; DB 23; Length 4557;

Best Local Similarity 66.8%; Pred. No. 2.5e-173; Mismatches 504; Indels 3; Gaps 1; Matches 1020; Conservative 0;

79 GGAGCCGAGCGCGGATGAAGGAGAGTCCAAAGATCGGCCCAAGACAGGAGGAGGAGG 138  
753 GGGAAACAAGACCAATGAAGAAAGTCCAAAATGCTCGCGGACTAGGAGGAGGAGG 812  
139 AAAATGGCGAGTTTACGAGCTTGCAGAGTGCCTCCGCTCGCTCGGCGCATCCTCCG 198  
813 AAAACAGCGAATTTATGAACCTGGCTTAATCTGCTTTGGCTTGGCTTATCAGCTCCG 872  
199 AGCTGGACAAGCGTCCATCCGCTCACACGAGCTACCTGATGAAGATCGCGCGCTCT 258  
873 AGTGGACAAGCATCCATATCAGACTCAGCAGCAGTATCTCAAAATGAGAGTGGTGT 932  
259 TCCCGGAGGTTTAGGAGACGCGTGGGAGCAGCGCGCGCGCGCGCGCGCGCGCGG 318  
933 TCCGAGAGGCGTCCGCGAGCGCTGGGCGCCACTCAAGTCGGACAGCGCGCGCGCGG 992  
319 TCGCAAGGAGCTGGGATCGCAGCTTGTGTCAGAGCTTTGGATGATTTGTTTGGTAG 378  
993 TTGCGCCAGAGACTGGGCTCCCATCTGCTCCAGACCTGGATGCTTATCTTCTGTTAG 1052  
379 CATCTGATGGCAAAATCATGTATATATCCAGAGCGCTTCTGTCATTTAGGCTTATCC 438  
1053 CCCAGATGGGAGATCATGTATCATCTCAGAGAGAGCGCTCAGTCCACTTGGTCTTCTC 1112  
439 AGTGGAGCTCAGCGCAACAGTATTTATGAATACATCCATCTTTCACACAGATGAGA 498  
1113 AGGTAGAGCTCAGCGGAACAGCATTTATGAATACATTCACCGCGCAGACGACGAGA 1172  
499 TGACCGCTGTCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 558  
1173 TGACCGCGGTGTCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1232  
559 TAGAGAGCTGTCTTCTTCTCGAATGAATGTCTTGGCGAAAGAACCGCGCGCTGA 618  
1233 TCGAGCGCTGTCTTCTTCTCGAATGAATGTCTTGGCGAAAGAACCGCGCGCTGA 1292  
619 CCGTCCAGCGGATACAAAGGTTCATCCACTCAGTGGCTACTTGAAGATCCGCGCATATGC 678  
1293 CCGTGGCGGCTACAAAGGTTCATCCACTCAGTGGCTACTTGAAGATCCGCGCATATGC 1352  
679 TGGACATGCTCTGTACGACTCTCTGCTACGAGTGTGGGCTGTGGCGCTGGGCCAGT 738  
1353 TGGACATGCTCTGTACGACTCTCTGCTACGAGTGTGGGCTGTGGCGCTGGGCCAGT 1412  
739 CGCTGCCACCGAGTGCATCCAGGATCAAGCTGTACAGTAACTGATGTTTCATGTTACGG 798  
1413 CGCTGCCACCGAGTGCATCCAGGATCAAGCTGTACAGTAACTGATGTTTCATGTTACGG 1472  
799 CCAGCTTGCCTGAGCTGATGATTTCTCGATTCAGGCTGACCGAGTGGCGGTATAGG 858  
1473 CCAGCTTGCCTGAGCTGATGATTTCTCGATTCAGGCTGACCGAGTGGCGGTATAGG 1532

1698 CTGAGTGCAGCCAGGTGGAACTGGAGAACTAGTCCCGCCACAAATGCTGTGGCTCC 1757  
1347 CCAGAACTGAGCCCGCCACTCAGAAAGAGTGAAGTCTGTACAGCGCATCTACAGCCTG 1406  
1758 CCGGAACAGCAGCTCCATTGAGAGCCAGCGACCTTTATATGGCGCGCCCTACAGCCTC 1817  
1407 CCCTTCTCTACCATATGAGACACTTCCCTCTGGACTTCTCACTTCTTACAGCAAAAG 1466  
1818 CCCTTCTCTACCATATGAGACACTTCCCTCTGGACTTCTCACTTCTTACAGCAAAAG 1877  
1467 CCAATGTTTCGGCGCAAGTTCGGCGAGCCCAAGGATCCCTCTGTGAGTGGCGACGCTT 1526  
1878 CCAGGATTCGCGCGCAAGTTCGGCGAGCCCAAGGATCCCTCTGTGAGTGGCGACGCTT 1937  
1527 TTCCTGAGCACAATGCCAGCGCGTGAATGCCAGTGGCATATGCAACCCCTAGT 1586  
1938 TTCCTGAGCACAATGCCAGCGCGTGAATGCCAGTGGCATATGCAACCTCTCTAGT 1997  
1587 CCTAGCAGCTCGTCTCCAGCTTAAATCCTCCAGAGC---CACCGGGAACACTGCTAGG 1643  
1998 CCCAGCAGCTATGCCAGCTTAAATCCTCCAGAGC---CACCGGGAACACTGCTAGG 2057  
1644 CACAGCTGCTGCGCAAGCTACGAG 1668  
2058 CATGGTCTGTGCGCAACCTATGAG 2082

RESULT 6

AAS92423 ID AAS92423 standard; cDNA; 4557 BP.

XX AAS92423;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #28227.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG28236.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 28227; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving















PF 17-JAN-2001; 2001WO-US01354.  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-02114886.  
 PR 30-JUN-2000; 2000US-02115135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216680.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 05-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 06-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 14-SEP-2000; 2000US-02321968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 25-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0234998.  
 PR 27-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 29-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0239345.  
 PR 13-OCT-2000; 2000US-0239345.  
 PR 20-OCT-2000; 2000US-0239377.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 20:59:29 ; Search time 181 Seconds  
(Without alignments)  
273.547 Million cell updates/sec

Title: US-09-923-684-12

Perfect score: 24

Sequence: 1 gagagcaagaagcacagcaagcc 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 693762

Minimum DB seq length: 0

Maximum DB seq length: 25

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq1:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq3:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	10	US-09-923-684-12
2	13.6	56.7	25	12	US-09-961-001-62
3	13.6	56.7	25	15	US-10-098-263B-19661
4	13.2	55.0	19	11	US-09-969-373-2893
5	13.2	55.0	25	12	US-09-992-665-108
6	13.2	55.0	25	15	US-10-098-263B-44300
7	13.2	55.0	25	15	US-10-098-263B-67339
8	13	54.2	22	15	US-10-121-393-13
9	13	54.2	22	15	US-10-098-263B-20997
10	13	54.2	25	15	US-10-098-263B-74151
11	12.8	53.3	20	12	US-09-104-750-32
12	12.8	53.3	25	12	US-09-754-853A-902
13	12.8	53.3	25	15	US-10-215-112-966
14	12.8	53.3	25	15	US-10-215-112-7212
15	12.8	53.3	25	15	US-10-098-263B-33827
16	12.6	52.5	20	12	US-09-771-933-112

c	17	12.6	52.5	21	12	US-09-771-933-154
	18	12.6	52.5	22	10	US-09-915-060-11
	19	12.6	52.5	23	11	US-09-915-060-45
	20	12.6	52.5	23	11	US-09-901-484A-143
	21	12.6	52.5	23	11	US-09-853-526-143
	22	12.6	52.5	25	11	US-09-827-998-1375
	23	12.6	52.5	25	11	US-09-827-998-1375
	24	12.6	52.5	25	11	US-09-827-998-1375
	25	12.6	52.5	25	11	US-09-827-998-1377
	26	12.6	52.5	25	11	US-09-827-998-1377
	27	12.6	52.5	25	11	US-09-827-998-1379
	28	12.6	52.5	25	11	US-09-827-998-1380
	29	12.6	52.5	25	15	US-09-827-998-1381
c	30	12.4	51.7	25	10	US-10-098-263B-80130
	31	12.4	51.7	25	10	US-09-866-108-14540
	32	12.4	51.7	25	10	US-09-866-108-14541
	33	12.4	51.7	25	10	US-09-866-108-14542
	34	12.4	51.7	25	10	US-09-866-108-14543
c	35	12.4	51.7	25	15	US-09-853-688-31
	36	12.2	50.8	17	11	US-10-098-263B-85390
	37	12.2	50.8	19	12	US-09-864-785-672
c	38	12.2	50.8	20	10	US-09-864-636A-914
	39	12.2	50.8	20	10	US-09-452-599-159
c	40	12.2	50.8	21	15	US-10-239-804-45
	41	12.2	50.8	23	11	US-09-263-959-1262
	42	12.2	50.8	23	12	US-09-864-636A-938
	43	12.2	50.8	24	11	US-09-817-014-9
	44	12.2	50.8	25	10	US-09-757-041-7
c	45	12.2	50.8	25	15	US-10-215-112-8929

## ALIGNMENTS

## RESULT 1

US-09-923-684-12  
; Sequence 12, Application US/09923684  
; Patent No. US20020081613A1  
; GENERAL INFORMATION:  
; APPLICANT: Narayanan, Ramaswamy  
; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER  
; FILE REFERENCE: 6818-24  
; CURRENT APPLICATION NUMBER: US/09/923.684  
; CURRENT FILING DATE: 2001-09-17  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 24  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-923-684-12

Query Match 100.0%; Score 24; DB 10; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GAGAGCAAGAAGCACAGCAAGCC 24  
Db 1 GAGAGCAAGAAGCACAGCAAGCC 24

## RESULT 2

US-09-961-001-62  
; Sequence 62, Application US/09961001  
; Publication No. US20030109466A1  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Susan M. Freier  
; TITLE OF INVENTION: ANTISENSE MODULATION OF KSR EXPRESSION  
; FILE REFERENCE: RTS-0280  
; CURRENT APPLICATION NUMBER: US/09/961.001  
; CURRENT FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 87



; TITLE OF INVENTION: Human Microarray  
; FILE REFERENCE: 3118.1  
; CURRENT APPLICATION NUMBER: US/10/098,263B  
; CURRENT FILING DATE: 2003-01-08  
; PRIOR APPLICATION NUMBER: 60/276,759  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 131066  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 67339  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-098-263B-20297

Query Match 55.0%; Score 13.2; DB 15; Length 25;  
Best Local Similarity 83.3%; Pred. No. 1.2e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CAGAAAGCAGCAGCAGC 23  
DB 24 CGAGAAAGCATAGAAAGC 7

## RESULT 8

US-10-121-393-13  
; Sequence 13, Application US/10121393  
; Publication No. US2003011563A1  
; GENERAL INFORMATION:  
; APPLICANT: Pals, Marie Salome Soares  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Balde, Aladjje  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES RELATING TO PAPAYA RIPENING  
; CURRENT APPLICATION NUMBER: US/10/121,393  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/283,008  
; PRIOR FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Amplification  
US-10-121-393-13

Query Match 54.2%; Score 13; DB 15; Length 22;  
Best Local Similarity 76.2%; Pred. No. 1.4e+04;  
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22  
DB 2 ACAGTAGAAGAACCATTCAG 22

## RESULT 9

US-10-098-263B-20297/c  
; Sequence 20297, Application US/10098263B  
; Publication No. US20030104410A1  
; GENERAL INFORMATION:  
; APPLICANT: Mittman, Michael  
; TITLE OF INVENTION: Human Microarray  
; FILE REFERENCE: 3118.1  
; CURRENT APPLICATION NUMBER: US/10/098,263B  
; CURRENT FILING DATE: 2003-01-08  
; PRIOR APPLICATION NUMBER: 60/276,759  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 131066  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 20297  
; LENGTH: 25

; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-098-263B-20297

Query Match 54.2%; Score 13; DB 15; Length 25;  
Best Local Similarity 76.2%; Pred. No. 1.5e+04;  
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22  
DB 22 AGAGCAAGTAGTCACGAAAG 2

## RESULT 10

US-10-098-263B-74151  
; Sequence 74151, Application US/10098263B  
; Publication No. US20030104410A1  
; GENERAL INFORMATION:  
; APPLICANT: Mittman, Michael  
; TITLE OF INVENTION: Human Microarray  
; FILE REFERENCE: 3118.1  
; CURRENT APPLICATION NUMBER: US/10/098,263B  
; CURRENT FILING DATE: 2003-01-08  
; PRIOR APPLICATION NUMBER: 60/276,759  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 131066  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 74151  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-098-263B-74151

Query Match 54.2%; Score 13; DB 15; Length 25;  
Best Local Similarity 76.2%; Pred. No. 1.5e+04;  
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22  
DB 3 AGAGCAAGACACACAGAGAG 23

## RESULT 11

US-09-104-750-32  
; Sequence 32, Application US/09104750  
; Publication No. US20030104364A1  
; GENERAL INFORMATION:  
; APPLICANT: Russell, John  
; APPLICANT: Colpitts, Tracey  
; TITLE OF INVENTION: Methods and Reagents Useful  
; TITLE OF INVENTION: for Detecting Disease of the Breast  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/104,750  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 6121.US.01  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L

REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6121.US.P1  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-104-750-32

Query Match 53.3%; Score 12.8; DB 12; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.7e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 CAAGAAGCAGCAGCAA 21  
DB 3 CAAGCAGCAGCTGC 18

RESULT 12  
US-09-754-853A-902  
Sequence 902, Application US/09754853A  
Publication No. US20030005491A1  
GENERAL INFORMATION:  
APPLICANT: Haug, Brian M.  
APPLICANT: Parnell, Laurence D.  
APPLICANT: Parsons, Jeremy D.  
APPLICANT: Wang, Ming Li  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With Soybean Cyst Nematode Resistance  
FILE REFERENCE: 38-10(15810)B  
CURRENT APPLICATION NUMBER: US/09/754,853A  
CURRENT FILING DATE: 2001-01-05  
PRIOR APPLICATION NUMBER: US 60/174,880  
PRIOR FILING DATE: 2000-01-07  
NUMBER OF SEQ ID NOS: 1119  
SEQ ID NO 902  
LENGTH: 25  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: 318013\_region\_A3\_453879\_15\_Reverse\_Primer\_Seq  
US-09-754-853A-902

Query Match 53.3%; Score 12.8; DB 12; Length 25;  
Best Local Similarity 87.5%; Pred. No. 1.8e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGACGAAGAAAGCACA 17  
DB 10 AGACGAAGAAAGGACA 25

RESULT 13  
US-10-215-112-966/c  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966  
Sequence 966, Application US/10215112  
Publication No. US20030082596A1  
GENERAL INFORMATION:  
APPLICANT: Michael Mittmann  
TITLE OF INVENTION: Method of Genetic Analysis of Probes:  
TITLE OF INVENTION: Test3  
FILE REFERENCE: 3119  
CURRENT APPLICATION NUMBER: US/10/215,112  
CURRENT FILING DATE: 2002-08-08  
NUMBER OF SEQ ID NOS: 14936  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 966  
LENGTH: 25  
TYPE: DNA

US-10-215-112-966

Mon Jul 21 15:46:54 2003

us-09-923-684-12.szlm25.rnpb

Page 5

Search completed: July 17, 2003, 21:50:23  
Job time : 184 secs

**THIS PAGE BLANK (USPTO)**





```
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-C10156-
120101-011-g03&t3=2001-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 13.
Location/Qualifiers
1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10156"
/dev_stage="Adult"
/note="Organ: colon.ins; Vector: puc18; Site_1: Smal;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
low stringency conditions."
BASE COUNT 93 a 143 c 93 g 81 t
ORIGIN
Query Match 11.5%; Score 327.4; DB 14; Length 410;
Best Local Similarity 96.0%; Pred. No. 8.9e-52;
Matches 389; Conservative 0; Mismatches 11; Indels 5; Gaps 5;
QY 1250 CCCCCCAGCAATACAGCTCATTCCAAATGGACAAACTGGAAATGCGGCCAGCTCGGAAA 1309
Db 7 CCCCCCAGCAGGATATAGCTCGTT-CAATGGACAAACTGGAAATGCGGCCAGCTCGGAAA 65
QY 1310 CTGGAGAGCCAGTCCCC-TCGAAGCGCTGC-TCCTCTCCAGACATGCGGCCCACTCA 1367
Db 66 CTGGAGATCCAGTCCCCCGTGCAGCGCGTGTCTCTCCAGAACTGCGGCCCACTCA 125
QY 1368 GAAAGCAGTGCCTCTGTACACGCCATCTACAGCTGCGCTTCTCTCTACCATTTATGA 1427
Db 126 GAAAGCAGTGCCTCTGTACACGCCATCTACAGCTGCGCTTCTCTCTACCATTTACGA 185
QY 1428 CACTTCCCTCTGGACTCTCACTTCTTCAGCAGCAAAAGCCAAATGTTGCCGGCCAGTTC 1487
Db 186 CACTTCCCTCTGGACTCTCACTTCTTCAGCAGCAAAAGCCAAATGTTGCCGGCCAGTTC 245
QY 1488 GGGCAGCCCAAGGATCCCTTGTGAGTGGCAGCGCTTTTCTTCAGCAGCAATGCGGCC 1547
Db 246 GGGCAGCCCTTGGATCCCTTGTGAGTGGCAGCGCTTTTCTTCAGCAGCAATGCGGCC 305
QY 1548 AGCGGTGAATGCCAGTGGC-ATTATGCCAACCCCTAGTGCCTAGC-AGCTGCTCTCAG 1605
Db 306 AGCGGTGAATGCCAGTGGCAGTGGCAATATATGCCAACCCCTAGTGCCTAGCAGCTC 365
QY 1606 CTAATAATCTCCAGAGCCACCGGCAACACTGCTAGGCACAGCC 1650
Db 366 CTAATAATCTCCAGAGCCACCGGCAACACTGCTAGGCACAGCC 410
RESULT 3
BM144061 332 bp mRNA linear EST 29-NOV-2001
LOCUS ADP/2.25a Differentiating glial CG-4 cells RDA library Rattus
DEFINITION norvegicus cDNA, mRNA sequence.
ACCESSION BM144061
VERSION BM144061.1 GI:17154128
```

```
JOURNAL MEDLINE COMMENT
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"
/dev_stage="stage 25"
/note="vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 136 a 133 c 139 g 144 t 1 others
ORIGIN
Query Match 13.5%; Score 386.6; DB 13; Length 553;
Best Local Similarity 83.2%; Pred. No. 6e-63;
Matches 440; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 304 GGCCTGTGACGCGTCCGACAGGAGCTGGGATCGCACTTCTCTCAGAGCTTTGGATGAT 363
Db 25 GGACCTTGACAACGTGGCAGGAACTGGGATCTCTCTTACAGACTCTTGGATGAT 84
QY 364 TTGTTTTTGTGTAGCATCTGTATGCAAAATCATGTATATATCCGAGACCGCTTCTGTGCC 423
Db 85 TTGTTTTTGTGTAGCTGTGACGCGAATCATGTATATATCCGAGACTGCTCCGCTCC 144
QY 424 ATTAGGCTTATCCAGGTGGAGCTCACGGGCAACAGTATTTATGAATACATCCATCTT 483
Db 145 ATCTGGGACTGTCCAGGTGGAACTGACAGGCAATAGTATATATGAATACATCTATCCGT 204
QY 484 CTGACCACTGATGATGACCGTGTCTCTCAGCGGCCACCGCGCTGCACCACTCCACCTGC 543
Db 205 CGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 264
QY 544 TCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
Db 265 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
QY 604 GGAACGCGGCTGACCTGACGCGGATACAAAGTATCCACTGCACTGCTACTTGAAGA 663
Db 325 GGAATCGGGGACTGACATGTAGTGATACAGAGTATGATGATGATGATGATGATGATGATGAT 384
QY 664 TCAGGCAGTATATGCTGGACATGCTCCCTGTACGACTCTGTACAGATGTTGGGGCTGG 723
Db 385 TAAGGCAGTACATCTGGACATTTCTGTACGACTCTGTACAGATGTTGGGGCTGG 444
QY 724 TGGCGGTGGCCAGTCCCTGCCACCCAGTCCCATCAGGATCAAGCTGTACAGTAACA 783
Db 445 TGGCGGTGGGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 504
QY 784 TGTTTCATGTTTCAGGCGGCGCTTGCCTGAAAGTGTATTTCTTGGATTC 832
Db 505 TGTTTCATGTTTCAGGCGGCGCTTGCCTGAAAGTGTATTTCTTGGATTC 553
RESULT 2
BQ325425 410 bp mRNA linear EST 17-MAY-2002
LOCUS PM1-C10156-120101-011-g03 C10156 Homo sapiens cDNA, mRNA sequence.
DEFINITION BQ325425
ACCESSION BQ325425
VERSION BQ325425.1 GI:20939694
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 410)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
```

TITLE



[illegible]

Email: [crapabs-r@mail.nih.gov](mailto:crapabs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
DNA Library Preparation: Michael Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (IKEN)  
DNA Library Arrayed by The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
Contact: [anadagysystemsbiology.org](mailto:anadagysystemsbiology.org)

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 46 Row: k Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF

```

FEATURES
source
1. .4125
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4823388"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"

```

[illegible]

Mon Jul 21 15:46:56 2003

```
QY 2805 TCACTGCACCTCAGCGTGGTAAACAGAGTGAGACTGTCTCAAAAAAAAAAAAAA 2859
|||||
Db 3792 CCACTGCACCTCAGCGTGGGCAACAGAGTGAGACTGTCTCAAAAAAAAAAAAAA 3846
|||||

RESULT 9
BM803650/c
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML2184 row: b column: 10
High quality sequence stop: 632.
FEATURES
source
1..990
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5520297"
/clone_lib="NIH_MGC_71"
/tissue_type="leionvosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT
170 a 277 c 268 g 275 t

Query Match 7.9%; Score 226.8; DB 14; Length 990;
Best Local Similarity 83.5%; Pred. No. 8.5e-33;
Matches 269; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 2539 AAGTGCATCAGGATTAATGTCGCCAGCGCGGCTCAGTGGCTCACACCTGTAATCCCA 2598
|||||
Db 353 AAGATTTCTTAAATAATTAATTAACCTTGGCTGGCGGGTGGCTCATGCTGTAATCCCA 294
|||||

QY 2599 GCACCTTTGGGAGGCCAAGGTGGCGGATCACCTGAGTGCAGGAGTTGCCAAGCCT-G 2657
Db 293 GCACCTTAGGAGGCCGAGGGGGTGGATCACCTGAGTGCAGGAGTTCCGAGACCCAGCTGG 234
|||||

QY 2658 CCACAGCTGAACCCCACTCTCCACTAAATAACAAAATTAAGTTGGCGATGGTGTGA 2717
|||||
Db 233 CCAACATGGTGAACCTCGTCTCTACTAAATAACAAAATTAAGTTGGCGATGGTGTGA 174
|||||

QY 2718 GCACCTGTAATCCAGCTTACTTCTGGAGGCTCAGATAGAGGATCACTTTGAACCCGGAGG 2777
Db 173 GCGCTGTAAATCCAGCTTACTCGGAGGCTCAGGAGGAGAAATCACTTGATCTCTGGAGG 114
|||||

QY 2778 TGGAGGTTGAGTGAAGTCAATCATCTGCTCCAGCTCGGCTTAACAGAGTGAGA 2837
|||||
Db 113 TGGAGGTTGAGTGAAGTCAATCATCTGCTCCAGCTCGGCTTAACAGAGTGAGA 54
|||||

QY 2838 CTGCTCAAAAAAAAAAAAAA 2859
|||||

Db 53 CTCTGTCTCAAAAAAAAAAAAAA 32

RESULT 10
AQ382985/c
LOCUS
DEFINITION
631 bp DNA linear GSS 21-MAY-1999
RPC111-139114.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-139114,
DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Other GSSs: RPC111-139114.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES
source
1..631
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7553197"
/db_xref="taxon:9606"
/clone="RPCI-11-139114"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT
151 a 158 c 154 g 168 t

Query Match 7.9%; Score 225.2; DB 17; Length 631;
Best Local Similarity 86.9%; Pred. No. 1.7e-32;
Matches 259; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 2563 CCAGCCGGGCTTCAGTGGCTCACACCTGTAATCCAGACCTTTGGGAGGCCAAGTGGGC 2622
|||||
Db 405 CTAGACCGGGCAGCGTGGCTCATGCTGTAATCCAGACCTTTGAGAGCGCGGAGCAGGT 346
|||||

QY 2623 GGATCACCCTGAGGTGAGGAGTTGGCACAAGCCT-GCCAAACAAGCTGAACCCCACTCC 2681
|||||
Db 345 GGATCACCCTGAGGTGAGGAGTTGGCACAAGCCTGAGCCAGCCTAGCCCACTGCTACTCT 286
|||||

QY 2682 ACTTAAATAACAAAATTAAGTTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2741
|||||
Db 285 ATTAAATAATTAATAATTAAGTGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 226
|||||

QY 2742 GAGGCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTGCAGTGACCTAAGATC 2801
|||||
Db 225 GAGGCTGAGGAGGAGAAATCGCTGAACCCGCGAGGTGGAGTTGCAGTGAGCCAGATC 166
|||||

QY 2802 ACATCACTGCACCTCCAGCTGGGTACACAGAGTGAGACTGTCTCAAAAAAAAAAAAAA 2859
|||||
```

[illegible]

```

FEATURES
seq frame
Location/Qualifiers
1..514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HBMSC_cr28C08"
/clone_lib="Human bone marrow stromal cells"
/sex="mixed"
/tissue_type="bone marrow stroma"
/dev_stage="mixed"
/lab_host="XLI-Blue MRF'/SOLR"
/note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI; mRNA made from human bone marrow stroma, cDNA made by

```

Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)<sub>18</sub> tail. The sequence tag for this library is ATACGGCGTC.

TAG-LIB=OI-H-DIO  
TAG-TISSUE=lung with fibrosis  
TAG\_SEQ=ATACGGCGTC\*



oligo-dT priming. Directionally cloned. Size-selected for average insert size >0.5 kb. Library constructed by Dr. Marian Young and Dr. Pamela Gehron Robey (NIDCR). Library supplied by Dr. Libin Jia (NHGRI)\*

BASE COUNT 121 a 139 c 119 g 135 t

ORIGIN

Query Match 7.8%; Score 224.2; DB 9; Length 514;  
Best Local Similarity 84.3%; Pred. No. 2.6e-32;  
Matches 264; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 2548 TCAGGATTATGTGCCCGGCGGCTCAGTGGCTCACACCTGTATATCCAGCACTTTGG 2607  
DB 331 TCAAAATTAACTGACTTGGTGGCGGAGTGGCTCACACCTGTATATCCAGCACTTTGG 272  
QY 2608 GAGGCCAAGGTGGCGGAGTCACTCAGGTCAGGAGTTTGGCAGACAGCCTG-CCACACAGC 2666  
DB 271 GAGGCCAAGGAGCGGATCACTCAGGTCAGGAGTTTCGAGACCACTGATCGGCATGG 212  
QY 2667 TGAACCCCATCTCCACTTAAATAACAAAATTAGTTGGGATGGTGGTGGAGCACTGTGA 2726  
DB 211 TGAACCCCGGTTCTTACTTAAATAACAAAATTAGCCGGCGTGGTGGCGCCCTGTA 152  
QY 2727 ATCCAGCTACTCTCGAGGCTCAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTG 2786  
DB 151 ATCCAGCTACTCAGGAGGCTGAGGAGGAGAACTCACTTGAACCCGGGAGGTGGAGTTG 92  
QY 2787 CAGTGAGCTAAGATCAGTCACTCCAGCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAA 2846  
DB 91 CAGTGAGCTGAGTGGTCCACTGCCTCCAGCTGGCGGAGGAGGAACTCCGCTC 32  
QY 2847 AAAAAAAAAAAAAA 2859  
DB 31 AAAAAAAAAAAAAA 19

RESULT 15  
BG697217

LOCUS 832 bp mRNA linear EST 07-MAY-2001

DEFINITION 602660451F1 NCI\_CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4803469 5', mRNA sequence.

ACCESSION BG697217

VERSION BG697217.1 GI:13963184

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: James Cleaver, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM10699 row: f column: 14  
High quality sequence stop: 694.  
Location/Qualifiers  
1..832  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4803469"  
/clone\_lib="NCI\_CGAP\_Skn3"  
/lab\_host="PH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life

FEATURES  
source

BASE COUNT 214 a 187 c 230 g 201 t

ORIGIN

Query Match 7.8%; Score 222.6; DB 12; Length 832;  
Best Local Similarity 84.0%; Pred. No. 5.3e-32;  
Matches 263; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 2548 TCAGGATTATGTGCCCGGCGGCTCAGTGGCTCACACCTGTATATCCAGCACTTTGG 2607  
DB 375 TCAAAATTAACTGACTTGGTGGCGGAGTGGCTCACACCTGTATATCCAGCACTTTGG 434  
QY 2608 GAGGCCAAGGTGGCGGAGTCACTCAGGTCAGGAGTTTGGCAGACAGCCTG-CCAACAAGC 2666  
DB 435 GAGGCCAAGGAGCGGATCACCTGAGGTCAGGAGTTTCGAGACCACTGATCGGCATGG 494  
QY 2667 TGAACCCCATCTCCACTTAAATAACAAAATTAGTTGGGATGGTGGTGGAGCACTGTGA 2726  
DB 495 TGAACCCCGGTTCTTACTTAAATAACAAAATTAGCCGGCGTGGTGGCGCTGTA 554  
QY 2727 ATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTG 2786  
DB 555 ATCCAGCTACTCAGGAGGCTGAGGAGGAGAACTCACTTGAACCCGGGAGGTGGAGTTG 614  
QY 2787 CAGTGAGCTAAGATCAGTCACTCCAGCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAA 2846  
DB 615 CAGTGAGCTGAGATCGTGCCTCCAGCTCCAGCTGGCGGAGGAGGAACTCCGCTC 674  
QY 2847 AAAAAAAAAAAAAA 2859  
DB 675 AAAAAAAAAAAAAA 687

Search completed: July 17, 2003, 20:59:18  
Job time : 3906 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:35:07 ; Search time 131 Seconds  
(without alignments)  
6693.045 Million cell updates/sec

Title: US-09-923-684-2  
Perfect score: 2859  
Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	220.6	7.7	70000	4	US-09-851-896-3
C 2	219.8	7.7	15977	4	US-09-608-285A-59
C 3	219.6	7.7	72928	3	US-09-009-913-1
C 4	218.2	7.6	53526	3	US-08-658-136-2
C 5	218.2	7.6	53577	3	US-08-658-136-1
C 6	217.8	7.6	36741	4	US-09-301-665-3
C 7	217.8	7.6	11282	4	US-09-754-250-3
C 8	217	7.6	8758	4	US-09-799-345-3
C 9	216.6	7.6	152331	3	US-09-128-155-16
C 10	215.8	7.5	80246	4	US-09-078-294-4
C 11	215.8	7.5	80595	4	US-09-078-294-3
C 12	215.2	7.5	8453	4	US-09-167-681-45
C 13	215.2	7.5	29629	4	US-09-729-995-3
C 14	213.8	7.5	2532	4	US-09-799-345-1
C 15	212.6	7.4	11531	1	US-08-068-945A-1
C 16	212.6	7.4	11531	1	US-08-442-806-1
C 17	212.6	7.4	21234	4	US-09-810-671-3
C 18	212	7.4	36741	4	US-09-301-665-3
C 19	211.6	7.4	282	1	US-08-133-628-8
C 20	211	7.4	99500	4	US-09-798-096-10
C 21	211	7.4	168575	4	US-09-426-290-1
C 22	210.6	7.4	55827	4	US-09-813-133A-3
C 23	210.4	7.4	3867	4	US-09-347-114A-81
C 24	210.4	7.4	14581	4	US-08-520-373D-4
C 25	210.4	7.4	22481	4	US-08-367-841A-43
C 26	210.4	7.4	22481	5	PCT-US95-07201-43
C 27	210.4	7.4	22484	4	US-09-875-223-2

C 28	210.4	7.4	43950	4	US-09-735-934A-3	Sequence 3, Appl
C 29	210.2	7.4	15977	4	US-09-608-285A-59	Sequence 59, Appl
C 30	209.6	7.3	14747	4	US-09-608-285A-42	Sequence 42, Appl
C 31	209.6	7.3	80246	4	US-09-078-294-4	Sequence 4, Appl
C 32	209.6	7.3	80595	4	US-09-078-294-3	Sequence 3, Appl
C 33	209.2	7.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 34	209.2	7.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 35	209.2	7.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 36	208.6	7.3	31571	1	US-08-323-443B-1	Sequence 1, Appl
C 37	208.6	7.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 38	208.6	7.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 39	208.6	7.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 40	208	7.3	81001	4	US-09-750-580-1	Sequence 1, Appl
C 41	207.8	7.3	2115	1	US-08-395-800A-7	Sequence 7, Appl
C 42	207.8	7.3	4823	2	US-08-457-254-5	Sequence 5, Appl
C 43	207.8	7.3	4823	2	US-08-484-257-20	Sequence 20, Appl
C 44	207.8	7.3	4823	3	US-08-999-927-5	Sequence 5, Appl
C 45	207.8	7.3	4823	4	US-08-461-819-5	Sequence 5, Appl

## ALIGNMENTS

## RESULT 1

US-09-851-896-3/c  
; Sequence 3, Application US/09851896  
; Patent No. 6410325  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Susan M. Freier  
; APPLICANT: Andrew T. Watt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP  
; TITLE OF INVENTION: EXPRESSION  
; FILE REFERENCE: RTS-0220  
; CURRENT APPLICATION NUMBER: US/09/851,896  
; CURRENT FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 89  
; SEQ ID NO 3  
; LENGTH: 70000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-851-896-3

Query Match 7.7%; Score 220.6; DB 4; Length 70000;  
Best Local Similarity 86.4%; Pred. No. 4.5e-44;  
Matches 255; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY	2563	CCAGCGCGGCTCAGTGGCTCACACCTGTATATCCAGCACCTTTGGGAGGCCAAGGTGGGC	2622
DB	12155	CTAGGCCAGGTGCAGTGGCTCACCGCTGCAATCCAGCACCTTTGGGAGGCCAAGGTGGGC	12096
QY	2623	GGATCAGCTGAGGTTCAGAGTGTTCGCAACAGCCT-GCCAAACAGCTGAACCCCATCTCC	2681
DB	12095	AGATCAGCTGAGGTTCAGAGTGTTCGCAACAGCCTGGCCAAATATGATGAACCCCATCTCT	12036
QY	2682	ACTAAATAACAAAATTTAGTGGCATGTGGTGGAGCAGCTGTATCCAGTACTCTG	2741
DB	12035	GCTAAATAACAAAATTTAGTGGCATGTGGTGGAGCAGCTGTATCCAGTACTCTG	11976
QY	2742	GAGCTGAGATAGGAGATCATTTGAACCCGGAGGTGGAGTTCAGTTCAGTTCAGTTCAGTTC	2801
DB	11975	GAGCTGAGGAGGAGATCATTTGAACCCGGAGGTGGAGTTCAGTTCAGTTCAGTTCAGTTC	11916
QY	2802	ACATCACTGCAGTCCAGCCTGGTGGTTCAGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	2856
DB	11915	CGCCACTGCAGTCCAGCCTGGTGGTTCAGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	11861

## RESULT 2

US-09-608-285A-59  
; Sequence 59, Application US/09608285A  
; Patent No. 6335013

GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
TITLE OF INVENTION: POLYPEPTIDES  
FILE REFERENCE: 28110/36570  
CURRENT APPLICATION NUMBER: US/09/608,285A  
CURRENT FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 09/583,231  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 09/557,800  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350,836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273,447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/244,444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/122,449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 59  
LENGTH: 15977  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: CD39-L4/L66 Gene Sequence  
NAME/KEY: CDS  
LOCATION: (245)..(461)  
NAME/KEY: CDS  
LOCATION: (1454)..(1533)  
NAME/KEY: CDS  
LOCATION: (2734)..(2877)  
NAME/KEY: CDS  
LOCATION: (4364)..(4439)  
NAME/KEY: CDS  
LOCATION: (4679)..(4714)  
NAME/KEY: CDS  
LOCATION: (5326)..(5414)  
NAME/KEY: CDS  
LOCATION: (5723)..(5802)  
NAME/KEY: CDS  
LOCATION: (6751)..(6812)  
NAME/KEY: CDS  
LOCATION: (7758)..(7859)  
NAME/KEY: CDS  
LOCATION: (8712)..(8852)  
NAME/KEY: CDS  
LOCATION: (9831)..(9887)  
NAME/KEY: CDS  
LOCATION: (11613)..(11728)  
NAME/KEY: CDS  
LOCATION: (13146)..(13691)  
NAME/KEY: CDS  
LOCATION: (15702)..(15839)  
NAME/KEY: misc.feature  
LOCATION: (14871)  
OTHER INFORMATION: n = a or c or g or t  
US-09-608-285A-59  
Query Match 7.7%; Score 219.8; DB 4; Length 15977;  
Best Local Similarity 84.1%; Pred. No. 4.2e-44;  
Matches 248; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
US-09-009-913-1  
Query Match 7.7%; Score 219.6; DB 3; Length 72928;  
Best Local Similarity 85.1%; Pred. No. 8.1e-44;  
Matches 257; Conservative 0; Mismatches 44; Indels 1; Gaps 1;  
US-09-009-913-1/c  
Sequence 1, Application US/09009913  
Patent No. 6087485  
GENERAL INFORMATION:  
APPLICANT: Axys Pharmaceuticals, Inc.  
TITLE OF INVENTION: Asthma Related Genes  
NUMBER OF SEQUENCES: 339  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bozicevic & Reed, LLP  
STREET: 285 Hamilton Ave, Suite 200  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,913  
FILING DATE: 21-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: SEQ-4P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3231  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 72928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-09-009-913-1



19660 ACTGCATCCAGCCTGGCGGACAGCAAGACTCTATCTCAAAAAA 19608

Db

RESULT 7

US-09-754-250-3/c

; Sequence 3, Application US/09754250

; Patent No. 6376225

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui et al

; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF

; FILE REFERENCE: CL001063

; CURRENT APPLICATION NUMBER: US/09/754,250

; CURRENT FILING DATE: 2001-01-05

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 111282

; TYPE: DNA

; ORGANISM: Human

; NAME/KEY: misc.feature

; LOCATION: (1)...(111282)

; OTHER INFORMATION: n = A,T,C or G

US-09-754-250-3

Query Match 7.6%; Score 217.8; DB 4; Length 111282;

Best Local Similarity 85.5%; Pred. No. 2.6e-43;

Matches 254; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

RESULT 8

US-09-799-345-3/c

; Sequence 3, Application US/09799345

; Patent No. 6323016

; GENERAL INFORMATION:

; APPLICANT: YE, Jane et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: CL001156

; CURRENT APPLICATION NUMBER: US/09/799,345

; CURRENT FILING DATE: 2001-03-06

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 8758

; TYPE: DNA

; ORGANISM: Human

US-09-799-345-3

Query Match 7.6%; Score 217; DB 4; Length 8758;

2538 GAAGGTGACATCAGGATTATGTGCCCGAGCGCGGCTCAGTGGCTCACACCTGTAATCCC 2597

1314 GAAGAGGAGAAAGAAACAAGAGAGCGCCAGGACAGTGGCTCACACCTGTAATCCC 1255

2598 AGCACTTTGGGAGCCCAAGGTGGCGGATCACCTGAGGTGAGGATTTGGGACCACT- 2656

1254 AGCACTTTGGGAGCCCAAGGTGGCGGATCACCTGAGGATTTGGGACCACT- 1195

2657 GCCAACAAGCTGAAACCCCATCTCCACTAAAATACAAAATTTAGTTGGGCATGGTGGT 2716

1194 GCCAAGCTGGCAAAACCCGCTCTACTAAAATACAAAATTTAGTTGGGCATGGTGGT 1135

2717 AGCACTTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTTGAACCCGGAG 2776

1134 GGTGCTGTATCCAGCTACTCTGGAGGCTGAGGAGGATCACTTGAACCCGGAG 1075

2777 GTGGAGTTGCACTGAGTAAGATCACTGCACTCCAGCTGGTGGTAAACCCGGAG 2836

1074 GCGGAGTGGCACTGAGTTACCCCACTGCACTCCAGCTGGTGGTAAACCCGGAG 1015

2837 ACTGCTCAAAAAA 2859

1014 ACTCCATCTCAAAAAA 992

RESULT 6

US-09-301-665-3/c

; Sequence 3, Application US/09301665

; Patent No. 6207876

; GENERAL INFORMATION:

; APPLICANT: KELLENS, RODNEY E.

; APPLICANT: DATTA, SURJIT K.

; APPLICANT: BLACKBURN, MICHAEL R.

; TITLE OF INVENTION: ADENOSINE DEAMINASE DEFICIENT TRANSGENIC MICE AND

; TITLE OF INVENTION: METHODS FOR THE USE THEREOF

; FILE REFERENCE: UTSH-243

; CURRENT APPLICATION NUMBER: US/09/301,665

; CURRENT FILING DATE: 1999-04-28

; EARLIER APPLICATION NUMBER: 60/083,408

; EARLIER FILING DATE: 1998-04-29

; EARLIER APPLICATION NUMBER: 60/083,370

; EARLIER FILING DATE: 1998-04-28

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 36741

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-301-665-3

Query Match 7.6%; Score 217.8; DB 4; Length 36741;

Best Local Similarity 84.0%; Pred. No. 1.7e-43;

Matches 246; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2567 GCCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCAAGTGGCGGAT 2626

19900 GCCAGGCACAGTGGCTGATGCTGTAATCCAGCACTGTGGGAGGCCAAGTGGGTGAT 19841

2627 CACCTGAGGTGAGGATTTGGGACAGCTGCCAACAAGCTGAAACCCCATCTCCACTAA 2686

19840 CACTTGGAGTCAGGAGTTCGAGACCAAGCTGGAACACATGTGAAACCCCATCTCCGCTAA 19781

2687 AAATACAAAATTTAGTTGGCATGGTGGTGGAGCACTGTATCCAGCTACTCTGGAGGC 2746

19780 AAATACAAAATTTAGTTGGCATGGTGGTGGAGCACTGTATCCAGCTACTCTGGAGGC 19721

2747 TGAGATAGAGATCACTTGAACCCGAGGTGGAGGTTCAGTGAGCTAAGATCAATCATC 2806

19720 TGAGGCGAGAGATCACTTGAACCCGAGGTGGAGGTTCAGTGAGCTAAGATCAATCATC 19661

2807 ACTGCACTCCAGCTGGGTAAACAGAGTGCAGACTGTCTCAAAAAA 2859

```
Best Local Similarity 81.2%; Pred. No. 1.6e-43;
Matches 264; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

QY 2536 GTGAAGTGCATCAGGATTATGTGCCAGGCGGGCTCAGTGGCTCACCTGTGTAATC 2595
Db 6356 GTGAATAGAGTAAGAATGACTTCCTTTAGGCCAGGACGGTGGCTCAGCCTATATC 6297
QY 2596 CCAGCACTTTGGGAGGCCAAGGTGGCGGATCACTCAGGTGAGGTTGGCACAAGCC 2655
Db 6296 CCAGCACTCTGGAAGCCCTAGGTAGGTGATCACTTTGAGGTGAGGTTTGAGACCAGCC 6237
QY 2656 T-GCCACAAGCTGAACCCCATCTCCACTAAATAATACAAAAATAGTTGGCATGGTG 2714
Db 6236 TGCCCAACATGGTGAACCCCATCTCTACTAAATAATACAAAAATAGTGGCATGGTG 6177
QY 2715 TGAGCACTGTATATCCAGCTACTCTGGAGGCTGAGATGAGGAGTCACTTGAACCCGG 2774
Db 6176 TGACCCCTATATCCAGCTACTCTAGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGG 6117
QY 2775 AGTGGAGGTGCGAGTGAAGTCAATCACTGCACTCCAGCTCCAGCTGGTAACAGAGTG 2834
Db 6116 AGTGGAGTTCAGTTAGCTAGGATCAAGCCAGTGCATCCAGCTGGGCAACAGAGCA 6057
QY 2835 AGACTCTCTCAAAAAAATAAAAAA 2859
Db 6056 AGAGAGACTCCTTATTAAAAA 6032

RESULT 9
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128.155
; EARLIER FILING DATE: 1998-08-03
; EARLIER FILING DATE: 1998-07-02
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 7.6%; Score 216.6; DB 3; Length 152331;
Best Local Similarity 84.0%; Pred. No. 5.7e-43;
Matches 268; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

QY 2542 GTGACATCAGGATTATGTGCCAGGCGGGCTCAGTGGCTCACCTGTGTAATCCAGCA 2601
Db 150967 GTGAAGTAAGAATTACTATATCCAGGCCAGGACACTGCCCTGTGTAATCCAGCA 151026
QY 2602 CTTTGGAGGCCAAGTGGCGGATCACTGAGGTGAGGTTGGCACAAGCTGCA 2660
Db 151027 CTTTGGAGGCCAAGTGGCGGATCACTGAGGTGAGGTTGCAAGCAACCTGTGCA 151086
QY 2661 ACAAGCTGAACCCCATCTCCACTAAAAATAACAAAAATAGTTGGCATGGTGGTGACA 2720
Db 151087 ACATGGTGAACCCCATCTCTACTAAAAATAC-AAATATAGTGGCATGGTGGTGACA 151145
QY 2721 CCTGTAAATCCAGCTACTCTGGAGGTGAGATGAGGAGTCACTTGAACCCGGAGGTGG 2780
Db 151145 CCTGTAAATCCAGCTACTCTGGAGGTGAGATGAGGAGTCACTTGAACCCGGAGGTGG 2780

Best Local Similarity 81.2%; Pred. No. 1.6e-43;
Matches 264; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

QY 2536 GTGAAGTGCATCAGGATTATGTGCCAGGCGGGCTCAGTGGCTCACCTGTGTAATC 2595
Db 6356 GTGAATAGAGTAAGAATGACTTCCTTTAGGCCAGGACGGTGGCTCAGCCTATATC 6297
QY 2596 CCAGCACTTTGGGAGGCCAAGGTGGCGGATCACTCAGGTGAGGTTGGCACAAGCC 2655
Db 6296 CCAGCACTCTGGAAGCCCTAGGTAGGTGATCACTTTGAGGTGAGGTTTGAGACCAGCC 6237
QY 2656 T-GCCACAAGCTGAACCCCATCTCCACTAAATAATACAAAAATAGTTGGCATGGTG 2714
Db 6236 TGCCCAACATGGTGAACCCCATCTCTACTAAATAATACAAAAATAGTGGCATGGTG 6177
QY 2715 TGAGCACTGTATATCCAGCTACTCTGGAGGCTGAGATGAGGAGTCACTTGAACCCGG 2774
Db 6176 TGACCCCTATATCCAGCTACTCTAGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGG 6117
QY 2775 AGTGGAGGTGCGAGTGAAGTCAATCACTGCACTCCAGCTCCAGCTGGTAACAGAGTG 2834
Db 6116 AGTGGAGTTCAGTTAGCTAGGATCAAGCCAGTGCATCCAGCTGGGCAACAGAGCA 6057
QY 2835 AGACTCTCTCAAAAAAATAAAAAA 2859
Db 6056 AGAGAGACTCCTTATTAAAAA 6032

RESULT 10
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078.294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contlg
US-09-078-294-4

Query Match 7.5%; Score 215.8; DB 4; Length 80246;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 2561 CCCAGGCCGGCTCAGTGGCTCACACCTGTATATCCAGCAAGCTTGGAGGCCAAGTGG 2620
Db 14331 CTCGGGCCAGGCATGGTGGCTTACGCTGTATATCCAGCAAGCTTGGAGGCCAAGTGG 14272
QY 2621 GCGGATCAGCTCAGTGGAGGTTGGCACAAGCTTGGCACAAGCTTGGCACAAGCTTGG 2679
Db 14271 GCGGATCAGCTCAGTGGAGGTTGGCACAAGCTTGGCACAAGCTTGGCACAAGCTTGG 14212
QY 2680 CCACATAAATAACAAAAATAGTTGGGCGATGGTGGGCGCTATATCCAGCTACTC 2739
Db 14211 CTACAAAAACACAAAAATAGTTGGGCGATGGTGGGCGCTATATCCAGCTACTC 14152
QY 2740 TGGAGGCTGAGATGAGGAGTCACTTGAACCCGGAGGTGGAGTTGCAGTGAGCTAAGA 2799
Db 14151 AGGAGGCTGAGGAGTCACTTGAACCCGGAGGTGGAGTTGCAGTGAGCTAAGA 14092
QY 2800 TCACATCACTCACTCCAGCTGGTAAACAGAGTGGTAAACAGAGTGGTAAACAGAGTGG 2854
Db 14091 TCAGTCACTCACTCCAGCTGGTAAACAGAGTGGTAAACAGAGTGGTAAACAGAGTGG 14037

RESULT 11
US-09-078-294-3/c
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078.294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
```

```
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3
Query Match          7.5%; Score 215.8; DB 4; Length 80595;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 2561 CCCAGGCGGGCTCAGTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCCAAGGTGG 2620
Db 14598 CTCTGGCCAGGCTGAGTGGCTTACGCTGTAAATCCAGCACTTTAGGAGGTCGAGGTGG 14539

QY 2621 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 2679
Db 14538 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 14479

QY 2680 CCACATAAAATACAAAAATTAGTTGGGATGTTGGTGCAGACCTGTAAATCCAGCACTCTC 2739
Db 14478 CTACAAAAACACAAAAATTAGTTGGGATGTTGGTGCAGACCTGTAAATCCAGCACTCTC 14419

QY 2740 TGGAGGCTGAGATAGGAGATCACTTGAACCCGGGAGGTGGAGTTGCAGTGAGCTAAGA 2799
Db 14418 AGGAGGCTGAGGATCAGAAATCGCTTGAACCCGGGAGGTGGAGTTGCAGTGAGCTGTGA 14359

QY 2800 TCACATCACTGCACCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA 2854
Db 14358 TCAGTCACTGCACCTCCAGCTAGGCAACAGTGAGACTGCATCTCAAAAAA 14304

RESULT 12
US-09-167-681-45/c
; Sequence 45, Application US/09167681A
; Patent No. 6265561
; GENERAL INFORMATION:
; APPLICANT: Weinshilboum, M.D., Richard M.
; APPLICANT: Raflogianis, Rebecca B.
; APPLICANT: Wood, Thomas C.
; APPLICANT: Oiterness, Diane M.
; TITLE OF INVENTION: SULFOTRANSFERASE SEQUENCE VARIANTS
; FILE REFERENCE: 07039/118001
; CURRENT APPLICATION NUMBER: US/09/167,681A
; CURRENT FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 8447
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (4361)...(4507)
; NAME/KEY: CDS
; LOCATION: (4612)...(4737)
; NAME/KEY: CDS
; LOCATION: (4827)...(4925)
; NAME/KEY: CDS
; LOCATION: (6322)...(6447)
; NAME/KEY: CDS
; LOCATION: (6543)...(6638)
; NAME/KEY: CDS
; LOCATION: (7137)...(7316)
; NAME/KEY: CDS
; LOCATION: (7439)...(7553)
US-09-167-681-45

Query Match          7.5%; Score 215.2; DB 4; Length 8453;
Best Local Similarity 85.1%; Pred. No. 4.4e-43;
Matches 252; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 2565 AGGCGGGCTCAGTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCCAAGGTGGCGG 2624
Db 2000 AGGCTGAGGCTGGCTCAGCTCAGCTGTAAATCCAGCACTTTGGGAGGTGCGATCAGCAGG 1941

QY 2625 ATCACTGAGGTGAGGAGTTTGGCA-CAAGCTGCCAACAGCTGAAACCCCATCTCCAC 2683

; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3
Query Match          7.5%; Score 215.8; DB 4; Length 80595;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 2561 CCCAGGCGGGCTCAGTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCCAAGGTGG 2620
Db 14598 CTCTGGCCAGGCTGAGTGGCTTACGCTGTAAATCCAGCACTTTAGGAGGTCGAGGTGG 14539

QY 2621 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 2679
Db 14538 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 14479

QY 2680 CCACATAAAATACAAAAATTAGTTGGGATGTTGGTGCAGACCTGTAAATCCAGCACTCTC 2739
Db 14478 CTACAAAAACACAAAAATTAGTTGGGATGTTGGTGCAGACCTGTAAATCCAGCACTCTC 14419

QY 2740 TGGAGGCTGAGATAGGAGATCACTTGAACCCGGGAGGTGGAGTTGCAGTGAGCTAAGA 2799
Db 14418 AGGAGGCTGAGGATCAGAAATCGCTTGAACCCGGGAGGTGGAGTTGCAGTGAGCTGTGA 14359

QY 2800 TCACATCACTGCACCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA 2854
Db 14358 TCAGTCACTGCACCTCCAGCTAGGCAACAGTGAGACTGCATCTCAAAAAA 14304

RESULT 13
US-09-729-995-3
; Sequence 3, Application US/09729995
; Patent No. 6426206
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00904
; CURRENT APPLICATION NUMBER: US/09/729,995
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 29629
; TYPE: DNA
; ORGANISM: Human
US-09-729-995-3

Query Match          7.5%; Score 215.2; DB 4; Length 29629;
Best Local Similarity 83.9%; Pred. No. 6.9e-43;
Matches 255; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 2557 TGTGCCCCAGCGCGGCTCAGTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCCAAG 2616
Db 11621 TCTGCTCCAGCTGGGCGGCTGCTCAGCGCTGTAAATCCAGCACTTTGGGAGGCCGAG 11680

QY 2617 GTGGCGGATCACTGAGTGCAGGAGTTTGCACAAGCCTG-CCACAAGCTGAAACCCC 2675
Db 11681 GTGGGTGATCACCTGAGTGCAGGAGTTCAAGACAGCCTGACCAACATGCGAAACCT 11740

QY 2676 ATCTCCACTAAATACAAAAATTAGTTGGGATGTTGGTGCAGCACCCTGTAAATCCAGCT 2735
Db 11741 GTCTCTACTAAAAATACAAAAATAGTACCGCGGCTGGTGGCAGACACCTGTAAATCCCACT 11800

QY 2736 ACTCTGGAGGCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTGCAGTGAGCT 2795
Db 11801 ACTCGGAGGCTGAGGAGGAGAAATCTCTTGNACCTGGGAGGTGGAGGCTGCAGTGAGCC 11860

QY 2796 AAGATCACATCACTGCACCTCCAGCCTGGGTACACAGATGAGACTGTCTCAAAAAA 2855
Db 11861 GAGATCACGCCATTGCACTCCAGCCTGGGAGGAGAGTGTCTCTCAAAAAA 11920

QY 2856 AAAA 2859
Db 11921 AAAA 11924

RESULT 14
US-09-799-345-1/c
; Sequence 1, Application US/09799345
; Patent No. 6323016
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
```



;; TITLE OF INVENTION: THEREOF  
;; FILE REFERENCE: CL001156  
;; CURRENT APPLICATION NUMBER: US/09/799,345  
;; CURRENT FILING DATE: 2001-03-06  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1  
;; LENGTH: 2532  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-799-345-1

Query Match 7.5%; Score 213.8; DB 4; Length 2532;  
Best Local Similarity 80.6%; Pred. No. 6.3e-43;  
Matches 262; Conservative 0; Mismatches 62; Indels 1; Gaps 1;  
  
QY 2536 GTGAAGGTGACATCAGGATTATGTGCCCCAGCGGGCTCAGTGGCTCACACCTGTAAATC 2595  
DB 1544 GTGAATAGAGATAAGATGACTTCTTTAGGCCAGGCAGCGTGGCTCACGCTTAAATC 1485  
  
QY 2596 CCAGCACCTTTGGAGGCCAAGGTGGCGGATCAGTGGCTCAGGAGTTCGACCAAGCC 2655  
DB 1484 CCAGCACCTCTGGAAGGCCAGGTAGTGGATCATTGAGGTGAGGATTTGAGACCAAGCC 1425  
  
QY 2656 T-GCCAAACAGCTGAACCCCATCTCCACTAAAAATACAAAAATAGTTGGGATGGTG 2714  
DB 1424 TGGCCAAATGTTGAAACCCCGTCTCTACTAAAAATACAAAAATAGCTGGGATGGTG 1365  
  
QY 2715 TGAGCACCTGTAATCCAGCTACTCTGGAGGCTCAGATAGGAGATCATTGAACCCGGG 2774  
DB 1364 TGCACGCTATAATCCAGCACCTCAGAGGCTGAGGAGGAGAAATGCTTGAACCCAGG 1305  
  
QY 2775 AGGTGGAGGTTCAGTGGAGTAAAGATCAGATCAGTCCAGCTGGGTAACAGAGTG 2834  
DB 1304 AGGTGGAGGTTCAGTGGAGTAAAGATCAGATCAGTCCAGCTGGGTAACAGAGTG 1245  
  
QY 2835 AGACTGCTCAAAAAAAGAAAAA 2859  
DB 1244 AGAGAGACTCTTATTAAAAA 1220

RESULT 15  
US-08-068-945A-1  
; Sequence 1, Application US/08068945A  
; Patent No. 5616483  
; GENERAL INFORMATION:  
; APPLICANT: Bjursell, Gunnar  
; APPLICANT: Carlsson, Peter  
; APPLICANT: Enerback, Sven  
; APPLICANT: Hansson, Lennart  
; APPLICANT: Lidberg, Ulf  
; APPLICANT: Nilsson, Jeanette  
; APPLICANT: Tornell, Jan  
; TITLE OF INVENTION: New DNA Sequences  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10036-2787  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/068,945A  
; FILING DATE: 27-MAY-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9201809-2

FILING DATE: 11-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9201826-6  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9202088-2  
FILING DATE: 03-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9300902-5  
FILING DATE: 19-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)819-8783  
TELEFAX: (212)354-8113  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11531 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Mammary gland  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1653..1727, 4071..4221, 4307..4429, 4707  
LOCATION: ..4504, 6193..6323, 6501..6608, 6751..6868, 8335  
LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11391)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: join(1722..1727, 4071..4221, 4307..4429, 4707  
LOCATION: ..4504, 6193..6323, 6501..6608, 6751..6868, 8335  
LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11391)  
OTHER INFORMATION: /EC\_number= 3.1.1.1  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..1640  
FEATURE:  
NAME/KEY: TATA\_signal  
LOCATION: 1611..1617  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1641..1727  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4071..4221  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4307..4429  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4707..4904  
FEATURE:  
NAME/KEY: exon  
LOCATION: 6193..6323  
FEATURE:  
NAME/KEY: exon  
LOCATION: 6501..6608  
FEATURE:  
NAME/KEY: exon  
LOCATION: 6751..6868  
FEATURE:  
NAME/KEY: exon  
LOCATION: 8335..8521  
FEATURE:  
NAME/KEY: exon  
LOCATION: 8719..8922  
FEATURE:

Mon Jul 21 15:46:55 2003

```

; NAME/KEY: exon
; LOCATION: 10124...10321
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10650...11490
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 11491...11531
; US-08-068-945A-1

Query Match      7.48; Score 212.6; DB 1; Length 11531;
Best Local Similarity 84.78; Pred. No. 2.1e-42; Indels 1; Gaps 1;
Matches 250; Conservative 0; Mismatches 44;

QY 2566 GGCCGGCTCAGTGGCTCACACCTGTAAATCCAGCACCTTTGGAGGCCAAGGTGGCGGGA 2625
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1971 GGCCGGCGCAGTGGCTCACACCTGTCAATCCACCAGTTTGGAGGCCGAGGTGGGTGA 2030
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2626 TCACCTGAGGTGAGGAGTTTGGCGACAGCCT-GCCACACAGCTGAACCCCATCTCCACT 2684
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2031 TCACTTGAGGTGAGGAGTTGGAGACCAGCCTGGGCCAATGGTGAACCCCGTCTCTACT 2090
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2685 AAAAATACAAAAATTAGTTGGGCATGGTGGTGCACCTGTATCCAGCTACTCTGGAG 2744
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2091 AAAAATACAAATATTAATCGGCGTGGTGGTGGTGCCTGTATCCAGCTACTCGGGAG 2150
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2745 GCTGAGATGAGGAGATCACTTTGAACCCGGGAGGTGGAGTTGCAGTGAGCTAAGATCACA 2804
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2151 GCTGAGGCAGGAGAACCGCTTGAAGCTGGGAGGTGGAGATTGGCGTGAGCTGAGATCAG 2210
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2805 TCACCTGCACCTCCAGCTGGTAAACAGAGTGAGACTGTCTCAAAAAAAGAAAAA 2859
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2211 CTACTGCACCTCAGCCTGGGTGACAGGCGGAGACTCTCTCAAAAAATAGAAAA 2265
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: July 17, 2003, 21:01:41  
Job time : 138 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:42:27 ; Search time 601 Seconds  
(without alignments)  
9813.842 Million cell

Title: US-09-923-684-2

perfect score: 2859  
Sequence: 1 actcactatagggtctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

[illegible]

searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

[illegible]

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
```

```
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
```

```
3: /cqn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cqn2_6/ptodata/2/pubpna/US06_PUBCOMP.seq:*
```

```
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
```

```
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
```

```
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB_seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB_seq:*
```

```

5: /cqn2_6/ptodata/z/pubpna/US08_PUBCOMB.seq:★
9: /cqn2_6/ptodata/2/pubpna/US08_NEW_PUBCOMB.seq:★

```

```
10: /cgn2_6/ptodata/2/pubpna/us09_new_pub_seq1:
```

11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:

```
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3:
```

```

13: /cqn2_3/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14: /cqn2_6/ptodata/2/pubpna/US10_NEW_PUB

```

```
15: /cgn2_6/ptodata/2/pubpna/us10_PUBCOMB.seq:*
```

```
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
```

```

17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2859	100.0		2859	10	US-09-923-684-2	Sequence 2, Appli
2	1661.6	58.1		3920	10	US-09-923-684-1	Sequence 1, Appli
3	465.6	16.3		472	10	US-09-923-684-10	Sequence 10, Appli
4	230.6	8.1		5000	12	US-09-927-214-6	Sequence 6, Appli
c	229.2	8.0		610	13	US-10-027-632-139070	Sequence 139070
5	228.6	8.0		897	15	US-10-027-632-121637	Sequence 121637
6	228.6	8.0		897	15	US-10-027-632-121638	Sequence 121638
7	228.6	7.9		489	15	US-10-027-632-287603	Sequence 287603
8	226.4	7.9		752	15	US-10-027-632-139071	Sequence 139071
9	225.8	7.9		586	15	US-10-027-632-264542	Sequence 264542
c	225.2	7.9		10612	11	US-09-764-877-3311	Sequence 3311, Ap
10	225.2	7.9		10614	11	US-09-764-877-3313	Sequence 3313, Ap
11	225.2	7.8		11337	11	US-09-764-877-2651	Sequence 2651, Ap
12	224.4	7.8		146547	12	US-10-017-128-1	Sequence 1, Appli
c	223.8	7.8		1743	12	US-09-764-891-6056	Sequence 6056, Ap
13	223.8	7.8		1745	12	US-09-764-891-6057	Sequence 6057, Ap

17	223	7.8	725	15	US-10-027-632-16541	Sequence 16541, A
18	223	7.8	725	15	US-10-027-632-16542	Sequence 16542, A
19	222.8	7.8	1812	15	US-10-037-632-100264	Sequence 100264, A
20	222.8	7.8	8207	11	US-09-764-877-2670	Sequence 2670, Ap
21	222.8	7.8	8207	12	US-09-764-877-741	Sequence 741, App
22	222.6	7.8	32134	12	US-09-764-891-6303	Sequence 6303, App
23	222.6	7.8	32134	12	US-10-025-428-608	Sequence 608, App
24	222.6	7.8	32191	12	US-09-764-891-6304	Sequence 6304, App
25	222.6	7.8	32191	15	US-10-025-428-609	Sequence 609, App
26	222.4	7.8	300000	15	US-10-262-552-33	Sequence 33, Appl
27	222.2	7.8	713	15	US-10-027-632-16630	Sequence 16630, A
28	222.2	7.8	1792	15	US-10-027-632-97316	Sequence 97316, A
29	222.2	7.8	1792	15	US-10-027-632-97317	Sequence 97317, A
30	221.4	7.7	624	15	US-10-027-632-97317	Sequence 110254, A
31	221.4	7.7	58837	11	US-09-982-091A-5	Sequence 5, Appl1
32	221.2	7.7	14781	11	US-09-764-877-2668	Sequence 2668, Ap
33	220.8	7.7	592	15	US-10-027-632-48886	Sequence 48886, A
34	220.8	7.7	32195	10	US-09-764-869-2017	Sequence 2017, Ap
35	220.8	7.7	32195	15	US-10-091-504-2017	Sequence 2017, Ap
36	220.8	7.7	32219	15	US-09-764-869-2016	Sequence 2016, Ap
37	220.8	7.7	32219	15	US-10-091-504-2016	Sequence 2016, Ap
38	220.6	7.7	119596	15	US-10-027-632-336-3	Sequence 3, Appl1
39	220.2	7.7	1537	15	US-10-027-632-263346	Sequence 263346, A
40	220	7.7	587	15	US-10-027-632-256765	Sequence 256765, A
41	220	7.7	587	15	US-10-027-632-256765	Sequence 256765, A
42	220	7.7	32249	12	US-09-764-891-7619	Sequence 258766, A
43	219.6	7.7	24699	11	US-09-764-877-2419	Sequence 7619, Ap
44	219.4	7.7	8369	10	US-09-764-903-67	Sequence 2419, Ap
45	219.4	7.7	8369	15	US-10-091-391-67	Sequence 67, Appl

## ALIGNMENTS

## RESULT 1

```

US-09-923-684-2
; Sequence 2, Application US/09923684
; Patent No. US20020081613A1
; GENERAL INFORMATION:
; APPLICANT: Narayanan, Ramaswamy
; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER
; FILE REFERENCE: 6818-24
; CURRENT APPLICATION NUMBER: US/09/923,684
; CURRENT FILING DATE: 2001-09-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-923-684-2

```

Query Match	100.0%;	Score 2859;	DB 10;	Length 2859;
Best Local Similarity	100.0%;	Pred. No. 0;		

Qy	1	ACTCACTATAGGGCTCGAGCGCGCCCGCGGCAGGTGGGCTCCGGGGCCCTGGAGCAGC	60
Db	1	ACTCACTATAGGGCTCGAGCGCGCCCGCGGCAGGTGGGCTCCGGGGCCCTGGAGCAGC	60
Qy	61	GC CGGGTCTAATATGCCCGGAGCGGCGCGATGAAGGAGAAGTCCAAAGATCGGGCCA	120
Db	61	GC CGGGTCTAATATGCCCGGAGCGGCGCGATGAAGGAGAAGTCCAAAGATCGGGCCA	120
Qy	121	AGACGAGGAGGAGAAGAAATGGCGAGTTTTACGAGCTTGCCAAAGTGTCTCCCGCTGC	180
Db	121	AGACGAGGAGGAGAAGAAATGGCGAGTTTTACGAGCTTGCCAAAGTGTCTCCCGCTGC	180
Qy	181	CGTCGGCCATCACATTCCAGCTCGACAAAGCGTCCATCATCGCGCTCACACAGAGCTACC	240
Db	181	CGTCGGCCATCACATTCCAGCTCGACAAAGCGTCCATCATCGCGCTCACACAGAGCTACC	240
Qy	241	TGAGATCGCGCGCTTCTCCCGAAGGTTTATGAGAGACGCGTGGGGACAGCCGAGCCGCG	300

Db	1321	GTCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCGAGCCCACTCAGAAAGCAGTGACC	1380
Qy	1381	TTCTGTACAGCATCTTACAGCCTGCGCTTCTCTCTACCATTTATGGACACTTCCCTCTGG	1440
Db	1381	TTCTGTACAGCATCTTACAGCCTGCGCTTCTCTCTACCATTTATGGACACTTCCCTCTGG	1440
Qy	1441	ACTCTCACTTCTTACAGCAAAAAAGCAATGTTCCCGCCCAAGTTCGGGAGGCCCAAG	1500
Db	1441	ACTCTCACTTCTTACAGCAAAAAAGCAATGTTCCCGCCCAAGTTCGGGAGGCCCAAG	1500
Qy	1501	GATCCCTTGTGAGTGGGACCGCTTTTCTCTGAGCAAAATGCGAGCCAGCGGTGAATGCC	1560
Db	1501	GATCCCTTGTGAGTGGGACCGCTTTTCTCTGAGCAAAATGCGAGCCAGCGGTGAATGCC	1560
Qy	1561	AGTGGCATTTATGCCAAACCCCTAGTGCCTAGCAGCTGCTCCAGCTTAAATCTCCAG	1620
Db	1561	AGTGGCATTTATGCCAAACCCCTAGTGCCTAGCAGCTGCTCCAGCTTAAATCTCCAG	1620
Qy	1621	AGCCACCGGGAACACTGCTAGGCACAGCTTGGTCCCAAGCTACGAAGTGGGTCAAGTCC	1680
Db	1621	AGCCACCGGGAACACTGCTAGGCACAGCTTGGTCCCAAGCTACGAAGTGGGTCAAGTCC	1680
Qy	1681	TGCTCTGGGGAAGTGGGAGGACTGCGACGCGCGGAGCGGAGCGGAGCGGAGTGC	1740
Db	1681	TGCTCTGGGGAAGTGGGAGGACTGCGACGCGCGGAGCGGAGCGGAGCGGAGTGC	1740
Qy	1741	GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTCCCTCTCTCTCTCTCTA	1800
Db	1741	GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTCCCTCTCTCTCTCTCTA	1800
Qy	1801	AGTAGGCTTGTCTGCTTCTTCTCAATGAGCTGCTCTCTGAGAGTGAAGAAATGG	1860
Db	1801	AGTAGGCTTGTCTGCTTCTTCTCAATGAGCTGCTCTCTGAGAGTGAAGAAATGG	1860
Qy	1861	CAGTCTGCTGCTCGGGGACACTAGTGCAGTATAAAGGCAAAAGGAGGAGTATC	1920
Db	1861	CAGTCTGCTGCTCGGGGACACTAGTGCAGTATAAAGGCAAAAGGAGGAGTATC	1920
Qy	1921	TGGCTTTCAGTAACTCTGCGCACTTACCAACCAAGGGGAGAGTATTTTCAAAA	1980
Db	1921	TGGCTTTCAGTAACTCTGCGCACTTACCAACCAAGGGGAGAGTATTTTCAAAA	1980
Qy	1981	CCAGTCCCATGTCTGAGAACACCCAGCTGCAATTTCTTTGCAAGATCTCTTCCACT	2040
Db	1981	CCAGTCCCATGTCTGAGAACACCCAGCTGCAATTTCTTTGCAAGATCTCTTCCACT	2040
Qy	2041	CCACCAAGTGAATTTTGGAGAACGCGCTTATTTGGCTATTTTCCCATGCCAGTTT	2100
Db	2041	CCACCAAGTGAATTTTGGAGAACGCGCTTATTTGGCTATTTTCCCATGCCAGTTT	2100
Qy	2101	GGAAGTGGGAAAACTATGTTGGAAATTTTGGGCTTGGGAGACAGAAATGCCACTACCA	2160
Db	2101	GGAAGTGGGAAAACTATGTTGGAAATTTTGGGCTTGGGAGACAGAAATGCCACTACCA	2160
Qy	2161	ACCCAGGCAAGAACAAACCCCTCAGGCTCAGTTCTTCCCTTCTTCTTCTTCTTCTT	2220
Db	2161	ACCCAGGCAAGAACAAACCCCTCAGGCTCAGTTCTTCCCTTCTTCTTCTTCTTCTT	2220
Qy	2221	AGCTGTGATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	2280
Db	2221	AGCTGTGATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	2280
Qy	2281	CTAGTTTCAACGAAAGAAATGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	2340
Db	2281	CTAGTTTCAACGAAAGAAATGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	2340
Qy	2341	TCGCGGAGATACCCAGCTTCCAGGACGAGACTGCTCTCTTCTTCTTCTTCTTCTT	2400
Db	2341	TCGCGGAGATACCCAGCTTCCAGGACGAGACTGCTCTCTTCTTCTTCTTCTTCTT	2400
Qy	2401	AACCCAGGCAAGAAAGTGAAGTGGGAAACAGAGGTTTAGCTGAGGACATGATTTCCA	2460
Db	2401	AACCCAGGCAAGAAAGTGAAGTGGGAAACAGAGGTTTAGCTGAGGACATGATTTCCA	2460

Db	241	TGAAGTGGCGCGCTTCTCCGGAAGTTTAGAGACGCGTGGGACAGCCGAGCGCG	300
Qy	301	CCGGGCGCTGGAGCGGCTGCCAAGAGCTGGATCGCATCTTCTGCTGAGACTTTGGATG	360
Db	301	CCGGGCGCTGGAGCGGCTGCCAAGAGCTGGATCGCATCTTCTGCTGAGACTTTGGATG	360
Qy	361	GATTTGTTTGTGTTGATGCTGATGCAAAATCATGTATATATATGCGAGACCGCTTCTG	420
Db	361	GATTTGTTTGTGTTGATGCTGATGCAAAATCATGTATATATATGCGAGACCGCTTCTG	420
Qy	421	TCCATTTAGCTTATCCAGTGGAGCTCAGCGGCAACAGTATTTATGAATATCATCATC	480
Db	421	TCCATTTAGCTTATCCAGTGGAGCTCAGCGGCAACAGTATTTATGAATATCATCATC	480
Qy	481	CTTCTGACCAAGATGAGATGACCGCTGCTCTCAGCGGCCACGAGCGCTGCACACACC	540
Db	481	CTTCTGACCAAGATGAGATGACCGCTGCTCTCAGCGGCCACGAGCGCTGCACACACC	540
Qy	541	TGCTCCAAAGATGATGAGATGAGAGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTT	600
Db	541	TGCTCCAAAGATGATGAGATGAGAGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTT	600
Qy	601	AAAGAAACCGGCGCTGACCTGACGCGGATCAAGGTCATCCACTGCGAGTGGCTACTTGA	660
Db	601	AAAGAAACCGGCGCTGACCTGACGCGGATCAAGGTCATCCACTGCGAGTGGCTACTTGA	660
Qy	661	AGATCAGGCAATATGCTGGACATGCTCCTGTAGGACTCCTGCTACAGATTTGGGGC	720
Db	661	AGATCAGGCAATATGCTGGACATGCTCCTGTAGGACTCCTGCTACAGATTTGGGGC	720
Qy	721	TGTTGGCGGTGGCGAGTGGTCCACCGAGTGCATCAGGAGTCAAGCTGTACAGTA	780
Db	721	TGTTGGCGGTGGCGAGTGGTCCACCGAGTGCATCAGGAGTCAAGCTGTACAGTA	780
Qy	781	ACATGTTGATGTTTACGGGCGAGCTTGCCTGAACTGATATTTCTGGATTTCCAGGTTGA	840
Db	781	ACATGTTGATGTTTACGGGCGAGCTTGCCTGAACTGATATTTCTGGATTTCCAGGTTGA	840
Qy	841	CCGAGTGGCGGTGACGAGCGGAGGCTGATCGAGAGACCTTATACCATCAGTGC	900
Db	841	CCGAGTGGCGGTGACGAGCGGAGGCTGATCGAGAGACCTTATACCATCAGTGC	900
Qy	901	ACGGCTGCGAGCTGTTCCACTCCGCTACGACACACCTCTCTGTTGGTGAAGGCGCAG	960
Db	901	ACGGCTGCGAGCTGTTCCACTCCGCTACGACACACCTCTCTGTTGGTGAAGGCGCAG	960
Qy	961	TCACCAAGTACTACCGGCTGCTGTCAGGCGGCGGCTGGGTGTTGGTGCAGAGCT	1020
Db	961	TCACCAAGTACTACCGGCTGCTGTCAGGCGGCGGCTGGGTGTTGGTGCAGAGCT	1020
Qy	1021	ACGCCACCGTGGTGCACACAGCGCTGCTCCCGGCGCCACTGATCGTGAAGTGAAT	1080
Db	1021	ACGCCACCGTGGTGCACACAGCGCTGCTCCCGGCGCCACTGATCGTGAAGTGAAT	1080
Qy	1081	ATGTACTACGGAGTTGAATACAGGAATTCAGCTGCTCCCTGGAGCAGGTTCCACTG	1140
Db	1081	ATGTACTACGGAGTTGAATACAGGAATTCAGCTGCTCCCTGGAGCAGGTTCCACTG	1140
Qy	1141	CCAAAGTCCAGGACTCTGAGGACCGCTTGTCTACCTCACAGAAACTAGGAAATAG	1200
Db	1141	CCAAAGTCCAGGACTCTGAGGACCGCTTGTCTACCTCACAGAAACTAGGAAATAG	1200
Qy	1201	TGAACCCAAATACCAAGATGAAGCAAAAGCTGAGAACAAACCTTACCCCGCAGC	1260
Db	1201	TGAACCCAAATACCAAGATGAAGCAAAAGCTGAGAACAAACCTTACCCCGCAGC	1260
Qy	1261	AATACAGCTATTCCTAATGAGCAAACTGGATCGGCGGAGCTGCGAACTGAGAGCCA	1320
Db	1261	AATACAGCTATTCCTAATGAGCAAACTGGATCGGCGGAGCTGCGAACTGAGAGCCA	1320
Qy	1321	GTCCCTGCAAGCGCTGCTGCTCTCAGAACTGCGAGCCCACTCAGAAAGCAGTGACC	1380

QY 2461 TTGACCGAGACCTAACTGCGCTCAGAGTGTAGACCGGAGATGGTGCAGATCCCTGCAGT 2520  
DB |||||  
DB 2461 TTGACCGAGACCTAACTGCGCTCAGAGTGTAGACCGGAGATGGTGCAGATCCCTGCAGT 2520  
QY 2521 GCCATTAAATGTGGGTGAAGGTGACATCAGGATTTATGTCGCCAGGCGGGCTCAGTGG 2580  
DB |||||  
DB 2521 GCCATTAAATGTGGGTGAAGGTGACATCAGGATTTATGTCGCCAGGCGGGCTCAGTGG 2580  
QY 2581 CTCACACCTGTATCCAGCAGCTTTGGGAGGCCAAGGTGGGGGATCACTTGAGTGCAGG 2640  
DB |||||  
DB 2581 CTCACACCTGTATCCAGCAGCTTTGGGAGGCCAAGGTGGGGGATCACTTGAGTGCAGG 2640  
QY 2641 AGTTTGCACAAGCTGTCACAAGCTGAAACCCCATCTCCACTAAATAACAAAATTA 2700  
DB |||||  
DB 2641 AGTTTGCACAAGCTGTCACAAGCTGAAACCCCATCTCCACTAAATAACAAAATTA 2700  
QY 2701 GTTGGCATGTTGGTGAAGCTGTAATCCAGCTTACTCTGGAGCTGAGATAGGAGAT 2760  
DB |||||  
DB 2701 GTTGGCATGTTGGTGAAGCTGTAATCCAGCTTACTCTGGAGCTGAGATAGGAGAT 2760  
QY 2761 CACTTGAACCCGGGAGGTGGAGCTTGCAGTGAAGTCAAGATCACATCACTGCACCTCCAGCC 2820  
DB |||||  
DB 2761 CACTTGAACCCGGGAGGTGGAGCTTGCAGTGAAGTCAAGATCACATCACTGCACCTCCAGCC 2820  
QY 2821 TGGGTAAACAGAGTGAAGCTGTCTCAAAAAA 2859  
DB |||||  
DB 2821 TGGGTAAACAGAGTGAAGCTGTCTCAAAAAA 2859

## RESULT 2

US-09-923-684-1

; Sequence 1, Application US/09923684

; Patent No.: US200200813A1

; GENERAL INFORMATION:

; APPLICANT: Narayanan, Ramaswamy

; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER

; FILE REFERENCE: 6818-24

; CURRENT APPLICATION NUMBER: US/09/923.684

; CURRENT FILING DATE: 2001-09-17

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1

; LENGTH: 3920

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-923-684-1

Query Match 58.1%; Score 1661.6; DB 10; Length 3920;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1664; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACTCACTATAGGGCTCGAGCGCGCGCGGGCAGGTGGGGCTCCGGGCGCTGGAGCAG 60  
DB 1 ACTCACTATAGGGCTCGAGCGCGCGCGGGCAGGTGGGGCTCCGGGCGCTGGAGCAG 60  
QY 61 GCCGGGTCTAATATGCCGAGCGCGGCGGATGAAGAGTCCAAAGATGGGCGCA 120  
DB 61 GCCGGGTCTAATATGCCGAGCGCGGCGGATGAAGAGTCCAAAGATGGGCGCA 120  
QY 121 AGACCGAGGAGGAGAAATGGCGAGTTTACCGAGCTTGCCAAAGCTGCTCCCGCTGC 180  
DB 121 AGACCGAGGAGGAGAAATGGCGAGTTTACCGAGCTTGCCAAAGCTGCTCCCGCTGC 180  
QY 181 CGTCCGCCATCACTTCGCGAGCTGGACAAAGCTTCATCATCGCTCCAGAGCTACC 240  
DB 181 CGTCCGCCATCACTTCGCGAGCTGGACAAAGCTTCATCATCGCTCCAGAGCTACC 240  
QY 241 TGAAGATCGCGCGCTCTTCCCGAAGTTTACGAGCGCGTGGGAGACCGCGCCGCG 300  
DB 241 TGAAGATCGCGCGCTCTTCCCGAAGTTTACGAGCGCGTGGGAGACCGCGCCGCG 300  
QY 301 CCGGGCCCTGACGCGCTCGCCAAAGGAGCTGGGATCGCACTTGTGTCAGACTTTGGATG 360

DB |||||  
QY 301 CCGGGCCCTGACGCGCTCGCCAAAGGAGCTGGGATCGCACTTGTGTCAGACTTTGGATG 360  
DB |||||  
DB 361 GATTGTTTTTGGTGTAGCATCTGATGGCAAAATCATGTATATATCCGAGACCGCTTCTG 420  
QY 361 GATTGTTTTTGGTGTAGCATCTGATGGCAAAATCATGTATATATCCGAGACCGCTTCTG 420  
DB |||||  
DB 421 TCCATTTAGGCTTATCCAGGTGAGCTCACGGGGCAACAGTATTTATGTAATACATCCATC 480  
QY 421 TCCATTTAGGCTTATCCAGGTGAGCTCACGGGGCAACAGTATTTATGTAATACATCCATC 480  
DB |||||  
QY 481 CTCTGACCACGATGAGATGACCGCTGCTCTCACGGCCACACGCGCTGACACCGCC 540  
DB |||||  
DB 481 CTCTGACCACGATGAGATGACCGCTGCTCTCACGGCCACACGCGCTGACACCGCC 540  
QY 541 TGTCTCAAGAGTATGAGATAGAGAGTCTGTTCTTCTTCTGCAATGAATGTGCTTGGCGA 600  
DB |||||  
DB 541 TGTCTCAAGAGTATGAGATAGAGAGTCTGTTCTTCTTCTGCAATGAATGTGCTTGGCGA 600  
QY 601 AAAGGAACCGGGCCTGACCTGCAGCGGATACAAAGGTCAATCCACTGCACTGCTACTTGA 660  
DB |||||  
DB 601 AAAGGAACCGGGCCTGACCTGCAGCGGATACAAAGGTCAATCCACTGCACTGCTACTTGA 660  
QY 661 AGATCAGGCAATATATGCTGGACATGTCCCTGTACGACTCTGCTGCTGCTGCTGCTG 720  
DB |||||  
DB 661 AGATCAGGCAATATATGCTGGACATGTCCCTGTACGACTCTGCTGCTGCTGCTGCTG 720  
QY 721 TGGTGGCCGTGGCGCAGTCCCTGCCACCGAGTGCATCAGCGAGATCAAGCTGTACAGTA 780  
DB |||||  
DB 721 TGGTGGCCGTGGCGCAGTCCCTGCCACCGAGTGCATCAGCGAGATCAAGCTGTACAGTA 780  
QY 781 ACATGTTCAATGTTAGGGCCAGCTTGACCTGTAAAGTGTATTTCTGATTTCCAGGGTGA 840  
DB |||||  
DB 781 ACATGTTCAATGTTAGGGCCAGCTTGACCTGTAAAGTGTATTTCTGATTTCCAGGGTGA 840  
QY 841 CCGAGGTGACGGTTTACGAGCGCGAGACCTGATCGAGAACACACCTTATACCATCAGTGC 900  
DB |||||  
DB 841 CCGAGGTGACGGTTTACGAGCGCGAGACCTGATCGAGAACACACCTTATACCATCAGTGC 900  
QY 901 ACGGCTGGAGCTGTTCACCTCCGCTACGCACACACCTCTCTTGTGTAAGGGCCAGG 960  
DB |||||  
DB 901 ACGGCTGGAGCTGTTCACCTCCGCTACGCACACACCTCTCTTGTGTAAGGGCCAGG 960  
QY 961 TCACCAACCAAGTACTACCGCTGTGTCACAGCGGGGGCTGGGTGTGGTGTGAGAGCT 1020  
DB |||||  
DB 961 TCACCAACCAAGTACTACCGCTGTGTCACAGCGGGGGCTGGGTGTGGTGTGAGAGCT 1020  
QY 1021 ACGCCACCGTGTGCACAAACAGCGCTCGTCCCGGCGCCACTCGATCGTGTGTCATTT 1080  
DB |||||  
DB 1021 ACGCCACCGTGTGCACAAACAGCGCTCGTCCCGGCGCCACTCGATCGTGTGTCATTT 1080  
QY 1081 ATGTACTCAGCGAGATTTGAATACAAAGAACTTCAAGTGTCCCTGGAGCAGGTGTCCACTG 1140  
DB |||||  
DB 1081 ATGTACTCAGCGAGATTTGAATACAAAGAACTTCAAGTGTCCCTGGAGCAGGTGTCCACTG 1140  
QY 1141 CCAAGTCCCAGGACTCTCGAGGACCGCTTGTCTACCTCACAAGAACTAGGAAATTTAG 1200  
DB |||||  
DB 1141 CCAAGTCCCAGGACTCTCGAGGACCGCTTGTCTACCTCACAAGAACTAGGAAATTTAG 1200  
QY 1201 TGAACCCAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTTACCCCGCCACAGC 1260  
DB |||||  
DB 1201 TGAACCCAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTTACCCCGCCACAGC 1260  
QY 1261 AATACAGCTCATTTCCAAATGACAAACTGGAATGCGGGCAGCTCGGAAACTGAGAGCCA 1320  
DB |||||  
DB 1261 AATACAGCTCATTTCCAAATGACAAACTGGAATGCGGGCAGCTCGGAAACTGAGAGCCA 1320  
QY 1321 GTCCCGCTGCAAGCGTGTCTCTCCAGAACTGCAGCCCGCTCAGAACAGCAGTGACC 1380  
DB |||||  
DB 1321 GTCCCGCTGCAAGCGTGTCTCTCCAGAACTGCAGCCCGCTCAGAACAGCAGTGACC 1380  
QY 1381 TTCTGTACAGCCATCTTACAGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 1440  
DB |||||

Db	1381	TTCTGTACAGCCATCTACAGCGTGCCTTCTCTACCATACGACACTTCCCTCTGG	1440
Qy	1441	ACTCTACTTCTCAGCAGCAAAAGCCAAATGTTGGCGGCAAGTTCTGGGCAAGCCCAAG	1500
Db	1441	ACTCTACGCTTCTCAGCAGCAAAAGCCAAATGTTGGCGGCAAGTTCTGGGCAAGCCCAAG	1500
Qy	1501	GATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCACAATGCCAGCAGCGGTGAATGCC	1560
Db	1501	GATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCACAATGCCAGCAGCGGTGAATGCC	1560
Qy	1561	AGTGCCATATGCCAACCCCTAGTGTCTAGCAGCTGTCTCCAGCTTAAATAATCTCCAG	1620
Db	1561	AGTGCCATATGCCAACCCCTAGTGTCTAGCAGCTGTCTCCAGCTTAAATAATCTCCAG	1620
Qy	1621	AGCCACCGGCAACACTCTAGGCAAGCGCTGTGCGCAAGCTACGAAG	1668
Db	1621	AGCCACCGGCAACACTCTAGGCAAGCGCTGTGCGCAAGCTACGAAG	1668
RESULT 3			
US-09-923-684-10			
; Sequence 10, Application US/09923684			
; Patent No. US20020081613A1			
; GENERAL INFORMATION:			
; APPLICANT: Narayanan, Ramaswamy			
; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER			
; FILE REFERENCE: 6818-24			
; CURRENT APPLICATION NUMBER: US/09/923,684			
; CURRENT FILING DATE: 2001-09-17			
; NUMBER OF SEQ ID NOS: 16			
; SOFTWARE: Patentin version 3.1			
; SEQ ID NO 10			
; LENGTH: 472			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-923-684-10			
Query Match			
Best Local Similarity 16.3%; Score 465.6; DB 10; Length 472;			
Matches 468; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
Qy	1158	TGGAGGACGCGCTTGTCTACCTACAGAAACTAGGAAATAGTGAACCCCAAAATACC	1217
Db	1	TGGAGGACGCGCTTGTCTACCTACAGAAACTAGGAAATAGTGAACCCCAAAATACC	60
Qy	1218	AGATGAGACAAAGCTGAGCAACAAACCTTACCCCCACAGCAATACAGCTCATTCCAA	1277
Db	61	AGATGAGACAAAGCTGAGCAACAAACCTTACCCCCACAGCAATACAGCTCATTCCAA	120
Qy	1278	ATGACAAACTGGAATGCGGCCAGCTCGGAAACTGGAGAGCCAGTCCCTCGAAGCGCT	1337
Db	121	ATGACAAACTGGAATGCGGCCAGCTCGGAAACTGGAGAGCCAGTCCCTCGAAGCGCT	180
Qy	1338	GCTGCTCCTCAGAACTGAGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCC	1397
Db	181	GCTGCTCCTCAGAACTGAGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCC	240
Qy	1398	TACAGCTGCGCTTCTCTACCAATATGAGACTTCCCTCTGAGCTCTCACTTCTTCAGC	1457
Db	241	TACAGCTGCGCTTCTCTACCAATATGAGACTTCCCTCTGAGCTCTCACTTCTTCAGC	300
Qy	1458	AGCAAAAGCAATGTTGCGGCCAAGTTGGGCGAGCCCAAGATCCCTCTGTGAGGTG	1517
Db	301	AGCAAAAGCAATGTTGCGGCCAAGTTGGGCGAGCCCAAGATCCCTCTGTGAGGTG	360
Qy	1518	GCAGCTTTTCTTCTGAGCACAATGCCAGCCAGCGGTGAATGCCAGTGGCATATGCGAAC	1577
Db	361	GCAGCTTTTCTTCTGAGCACAATGCCAGCCAGCGGTGAATGCCAGTGGCATATGCGAAC	420
Qy	1578	CCCTAGTGCCTAGCAGCTCTCTCAGCTTAAATAATCTCCAGAGCCACGG	1629
Db	421	CCCTAGTGCCTAGCAGCTCTCTCAGCTTAAATAATCTCCAGAGCCACGG	472
RESULT 4			
US-09-902-214-6			
; Sequence 6, Application US/09902214			
; Publication No. US20030104521A1			
; GENERAL INFORMATION:			
; APPLICANT: Whittaker, Paul Andrew			
; TITLE OF INVENTION: Disease-Associated Gene			
; FILE REFERENCE: 4-31503A/H031			
; CURRENT APPLICATION NUMBER: US/09/902,214			
; CURRENT FILING DATE: 2001-07-10			
; NUMBER OF SEQ ID NOS: 84			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 6			
; LENGTH: 50000			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-902-214-6			
Query Match			
Best Local Similarity 8.1%; Score 230.6; DB 12; Length 50000;			
Matches 283; Conservative 0; Mismatches 69; Indels 1; Gaps 1;			
Qy	2508	AGATGCTCAGTGCCTATTAATGTTGGTGAAGGTGACATCAGGATTATGTGCCCCAGG	2567
Db	40656	AAATGCCAGAAAGAACTGCTAAAGGTGTGCACAGTGCCTCAAGAAGTATGTGACTGG	40715
Qy	2568	CGGGCTCAGTGCCTCAGACCTCTAATCCAGCACCTTTGGGAGGCCAAGTGGCGGATC	2627
Db	40716	CGGGCGCAGTGCCTCAGCGCTGTATCCAGCACCTTTGGGAGGCCGAGTGGCGAGTGC	40775
Qy	2628	ACCTGAGTCCAGGATTGGCACAAGCCT-GCCAACAAGCTGAAACCCCATCTCCACTAA	2686
Db	40776	ACCTGAGTTCAGAGTTGAGACCCAGCTGGGCAACATGACAAACCCCGTCTCTACTAA	40835
Qy	2687	AAATACAAAATTTAGTTGGGCATGGTGTGAGCACCTGTATCCAGCTACTCTGAGGC	2746
Db	40836	AGATACAAAATTTATCTGGGTGTGTGTGACCTGTATCCAGCTACTCTGAGGC	40895
Qy	2747	TGAGTAGGAGGATCATTGAACCCGGAGGTGGAGTTCAGTGAAGATCAGATC	2806
Db	40896	TGAGGCAGAAAGTTCGTTGAACCCGAGGAGTGGAGTTCAGTGAAGATCAGATC	40955
Qy	2807	ACTGACTCCAGCCTGGGTAAACAGAGTGAGACTCTCTCAAAAAAATAAAAA	2859
Db	40956	ACTGACTCCAGCCTGGGCGACAGCGAGACTCTGTCTCAAAAAAATAAAAA	41008
RESULT 5			
US-10-027-632-139070/c			
; Sequence 139070, Application US/10027632			
; GENERAL INFORMATION:			
; APPLICANT: Wang, David G.			
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide			
; TITLE OF INVENTION: Polymorphisms in the Human Genome			
; FILE REFERENCE: 108827.129			
; CURRENT APPLICATION NUMBER: US/10/027,632			
; CURRENT FILING DATE: 2002-04-30			
; PRIOR APPLICATION NUMBER: US 60/218,006			
; PRIOR FILING DATE: 2000-07-12			
; PRIOR APPLICATION NUMBER: US 60/198,676			
; PRIOR FILING DATE: 2000-04-20			
; PRIOR APPLICATION NUMBER: US 60/193,483			
; PRIOR FILING DATE: 2000-03-29			
; PRIOR APPLICATION NUMBER: US 60/185,218			
; PRIOR FILING DATE: 2000-02-24			
; PRIOR APPLICATION NUMBER: US 60/167,363			
; PRIOR FILING DATE: 1999-11-23			
; PRIOR APPLICATION NUMBER: US 60/156,358			
; PRIOR FILING DATE: 1999-09-28			
; PRIOR APPLICATION NUMBER: US 60/146,002			
; PRIOR FILING DATE: 1999-08-09			
; NUMBER OF SEQ ID NOS: 325720			



APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 139071  
LENGTH: 752  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-139071

Query Match 7.9%; Score 226.4; DB 15; Length 752;  
Best Local Similarity 87.5%; Pred. No. 5.7e-53;  
Matches 259; Conservative 0; Mismatches 36; Indels 1; Gaps 1;

QY 2564 CAGCCGGGCTCAGTGGCTCACACCTGTATATCCAGCAGCTTTGGGAGGCCAAGGTGGCG 2623  
DB 102 CTGGCCGGCGCGGTGGCTCATGCTGTATATCCAGCAGCTTTGGGAGGCCAAGGTGGCG 161  
QY 2624 GATCAGCTGAGTCAAGGATTTTGGCAAGCCT-GCCAAAGCTGAAACCCCATCTCCA 2682  
DB 162 GATCAGCTGAGTCAAGGATTTTGGCAAGCCTGAGCAGCTGGCCAAACATGTTGAAACCCCATCTCA 221  
QY 2683 CTAATAATACAAAAATTTAGTTGGCATGTTGGTGGAGCAGCTGTAAATCCAGCTACTCTGG 2742  
DB 222 CTAATAATACAAAAATTTAGTTGGCATGTTGGTGGAGCAGCTGTAAATCCAGCTACTCTGG 281  
QY 2743 AGGCTGAGTCAAGGATTTTGGCAAGCCTGAGCAGCTGGAGGTTGAGTGGAGTAAAGATCA 2802  
DB 282 AGGCTGAGTCAAGGATTTTGGCAAGCCTGAGCAGCTGGAGGTTGAGTGGAGTAAAGATCA 341  
QY 2803 CATCACTGCACTCCAGCCTGGGTAAACAGAGTGAGAGTGTCTCAAAAAAAGAAAAA 2858  
DB 342 CACCACTGCACTCCAGCCTGGGTAAACAGAGTGAGAGTGTCTCAAAAAAAGAAAAA 397

RESULT 10  
US-10-027-632-264542  
Sequence 264542, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23

524 GTTCAGTGGCCAGATCACACCTGCACTCCAGCCTGGGTGACAGAGTGAGACTCG 583  
QY 2843 TCAAAAAAAGAAAAA 2859  
DB 584 TCTCAAAAAATAAATA 600

RESULT 8  
US-10-027-632-287603  
Sequence 287603, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 287603  
LENGTH: 489  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-287603

Query Match 7.9%; Score 226.6; DB 15; Length 489;  
Best Local Similarity 82.6%; Pred. No. 4.2e-53;  
Matches 270; Conservative 1; Mismatches 55; Indels 1; Gaps 1;

QY 2534 GGGTGAAGTGACATCAGGATTTATGCCCCAGCGCGGTGAGTGGCTCACACCTGTAA 2593  
DB 148 GGGGAAAGTCCCATCAGAAACATCTAACACAGCGCGGCGGTGCTCATGCTGTAA 207  
QY 2594 TCCAGCAGCTTTGGGAGGCCAAGTGGCGGATCACCTGAGTCAAGGATTTGCGACAAG 2653  
DB 208 TCCAGCAGCTTTGGGAGGCCAAGTGGCGGATCACCTGAGTCAAGGATTTGCGACAAG 267  
QY 2654 CCT-GCCAAAGCTGAAACCCCATCTCCACTAAATAACAAAAATTTAGTTGGCATGTT 2712  
DB 268 CTGCGCAACATGTTGAAACCCCGTCTACTATAATTAACAAAAATTTAGTTGGCATGTT 327  
QY 2713 GGTGAGCAGCTGTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTTTGAACCCG 2772  
DB 328 GCGGCGGCTGTAATCCAGCTACTCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 387  
QY 2773 GGAGTGAGGTTGAGTGAAGTAAATCATCATCTGCACTCCAGCCTGGTAAACAGAG 2832  
DB 388 GGAGCGGAGGTTGAGTGAAGTAAATCATCATCTGCACTCCAGCCTGGTAAACAGAG 447  
QY 2833 TGAGACTGCTCAAAAAAAGAAAAA 2859  
DB 448 CAAGACTGCTCTCAAAAAAAGAAAAA 474

RESULT 9  
US-10-027-632-139071  
Sequence 139071, Application US/10027632  
GENERAL INFORMATION:



Query Match	7.9%	Score 225.2	DB 11	Length 10612
Best Local Similarity	85.4%	Pred. No. 3.4e-52		
Matches 251; Conservative	0	Mismatches 43	Indels 0	Gaps 0
QY	2566	GGCCGGGCTCAGTGCCTCACACCTGTAAATCCAGCACATT	TGGAGGCCAAGGTGGCGGA	2635
Db	2259	GGCTGGCGCAGTGCCTCAGCCCTGTATCCAGCACATT	TGGAGGCCAAGGCGGGTGA	2200
QY	2626	TCACCTGAGCTCAGGAGTTTGGCACAAGCTGCCACAACAGCT	GAACACCCCACTCTCCACTA	2685
Db	2199	TGCTTGAGGTCAAGAATTGCGATCAGCCTGGCCAAAAGGT	GAACCCCTTTCCTACTA	2140
QY	2686	AAATACAAAAAATTAGTTGGGCATGTTGGTGAGCACCTGTAT	TCCAGCTACTCTGGAGG	2745
Db	2139	AAATACAAAAAATTAGCTGGCCCTGGTGGTGCCTGTAAAT	TTTCAGCTACTCAGAGG	2080
QY	2746	CTGAGATAGGAGGATCACTTGAACCCGGAGGTGGAGGTTGC	AGCTAAGATCATCAT	2805

Query Match 7.8%; Score 224.4; DB 11; Length 11337;  
Best Local Similarity 87.4%; Pred. No. 5.8e-52;

Query Match 7.8%; Score 224.4; DB 11; Length 11337;  
Best Local Similarity 87.4%; Pred. No. 5.8e-52;

Query Match 7.8%; Score 224.4; DB 11; Length 11337;  
Best Local Similarity 87.4%; Pred. No. 5.8e-52;

Mon Jul 21 15:46:56 2003

us-09-923-684-2.rnpb

Matches 257; Conservative 0; Mismatches 36; Indels 1; Gaps 1;  
QY 2567 GCCGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGGGGAT 2626  
Db 2315 GCCAGGCGGTGGCTCATGCTGTAATCCAGGACTTTGGGGCGCGAGGTGGGTGGAT 2374  
QY 2627 CACCTGAGGTGAGGAGTTGGGACAGCCCT-GCCAAACAGCTGAACCCCTCTCCACTA 2685  
Db 2375 CACCTGAGGTGAGGAGTTGGGACAGCCCTGGCCCAACATGTTGAACCCCTCTCTACTA 2434  
QY 2686 AAAATACAAAATAGTTGGGATGGTGGTGGAGCAGCTGTAATCCAGCTACTCTGGAGG 2745  
Db 2435 AAAATGCAAAAATAGCGCGGCTGGTGGTGCACACCTGTAGTCCAGCTACTCAGGAGG 2494  
QY 2746 CTGAGATAGGAGATCATTTGAACCCGGGAGGTGGAGTTGCCAGTGAAGTAAATCAT 2805  
Db 2495 CTGAGGAGGAGATCATTTGAACCCGGGAGGTGGAGTTGCCAGTGAAGTAAATCAT 2554  
QY 2806 CACTGCAGTCCAGCTGGTGAACAGAGTGAGCTGTCTCAAAAAA 2859  
Db 2555 CACTGCAGTCCAGCTGGTGAACAGAGTGAGCTGTCTCAAAAAA 2608

RESULT 14  
US-10-017-128-1/c  
; Sequence 1, Application US/10017128  
; Publication No. US2003012436A1  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Jeanette  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT FOR VASCULAR DISEASE  
; FILE REFERENCE: MMI-001  
; CURRENT APPLICATION NUMBER: US/10/017,128  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/306,941  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/315,572  
; PRIOR FILING DATE: 2001-08-28  
; PRIOR APPLICATION NUMBER: US 60/327,488  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-017-128-1

Query Match 7.8%; Score 224.2; DB 15; Length 146547;  
Best Local Similarity 84.3%; Pred. No. 1.7e-51;  
Matches 264; Conservative 0; Mismatches 48; Indels 1; Gaps 1;  
QY 2548 TCAGGATTATGTGCCCCAGCGGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTGG 2607  
Db 75123 TCAAAATTAACGTGACTTGCTGGCGCAGTGGCTCACACCTGTAATCCAGCAGCTTGG 75064  
QY 2608 GAGCCCAAGGTGGCGGATCACCTGAGGTGAGGAGTTGGGACAAAGCTTG 2666  
Db 75063 GAGCCCAAGGTGGCGGATCACCTGAGGTGAGGAGTTGGGACAAAGCTTG 75004  
QY 2667 TGAACCCCTCTCCACTTAAATACAAAATAGTTGGGATGGTGGTGGAGCAGCTGTA 2726  
Db 75003 TGAACCCCTCTCTCTAAATACAAAATAGTTGGGATGGTGGTGGTGGTGGTGGTGA 74944  
QY 2727 ATCCAGCTACTCTGAGGTGAGATAGGAGGATCATTTGAACCCGGGAGGTGGAGTTG 2786  
Db 74943 ATCCAGCTACTCTGAGGTGAGGAGGATCATTTGAACCCGGGAGGTGGAGTTG 74884  
QY 2787 CAGTGAGTATAGATCATCATCTCCAGCTGGTGAACAGAGTGAGACTGTCTCAA 2846  
Db 74883 CAGTGAGTATAGATCATCTCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 74824  
QY 2847 AAAAAAAAAAAAAA 2859

Db 74823 AAAAAAAAAAAAAA 74811  
RESULT 15  
US-09-764-891-6056/c  
; Sequence 6056, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6056  
; LENGTH: 1743  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6056  
Query Match 7.8%; Score 223.8; DB 12; Length 1743;  
Best Local Similarity 87.1%; Pred. No. 4.2e-52;  
Matches 257; Conservative 0; Mismatches 37; Indels 1; Gaps 1;  
QY 2566 GGCGGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGCGGA 2625  
Db 409 GGCTGGTGTGTGGCTCACACCTATAATCCAGCAGCTTTGGGAGGCAAGGTGGCGGA 350  
QY 2626 TCACCTGAGTTCAGGAGTTTGGGACAAAGCTT-GCCAAACAGCTGAACCCCTCTCCACT 2684  
Db 349 TCACCTGAGTTCAGGAGTTTCAAGACAGCTTGGCCAAACATGGTGAACCCCTCTCTACT 290  
QY 2685 AAAAAATACAAAATTAAGTGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2744  
Db 289 AAAAAATACAAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 230  
QY 2745 GCTGAGATAGGAGTACACTTGAACCCGGGAGGTGGAGTTGTCAGTGGAGTGAAGTACCA 2804  
Db 229 GCTGAGGAGGAGGAGTGAAGTGAACCTGGGAGGTGGAGTTGTCAGTGGAGTGGAGTGG 170  
QY 2805 TCACCTGCACTCCAGCTGGGTAACAGAGTGAGACTGTCTCAAAAAA 2859  
Db 169 CCAGTGCAGCTCCAGCTGGGTGACAGAGTGAGACTGTCTTAAAAA 115

Search completed: July 17, 2003, 21:11:56  
Job time : 607 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:52:52 ; Search time 1857 Seconds  
(without alignments)  
376.127 Million cell updates/sec

Title: US-09-923-684-12  
Perfect score: 24  
Sequence: 1 gagagcaagaagacacagcaagcc 24

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 521186

Minimum DB seq length: 0  
Maximum DB seq length: 25

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	66.7	25	6	AX117664	AX117664 Sequence
2	15.2	63.3	21	6	AX452262	AX452262 Sequence
3	15.2	63.3	20	6	AX288411	AX288411 Sequence
4	15	62.5	24	6	AX293044	AX293044 Sequence
5	14.4	60.0	25	6	AX429796	AX429796 Sequence
6	14.2	59.2	20	6	AX191307	AX191307 Sequence
7	13.8	57.5	17	6	AX475389	AX475389 Sequence
8	13.8	57.5	25	6	AX476366	AX476366 Sequence
9	13.8	57.5	25	6	AX476367	AX476367 Sequence
10	13.8	57.5	25	6	AX476368	AX476368 Sequence
11	13.8	57.5	25	6	AX476369	AX476369 Sequence
12	13.8	57.5	25	6	AX476370	AX476370 Sequence
13	13.8	57.5	25	6	AX476371	AX476371 Sequence
14	13.8	57.5	25	6	AX476372	AX476372 Sequence
15	13.8	57.5	25	6	AX476373	AX476373 Sequence
16	13.8	57.5	25	6	AX476374	AX476374 Sequence
17	13.6	56.7	22	6	AR207807	AR207807 Sequence
18	13.4	55.8	20	4	BOVDIK08	D44509 Bovine DNA,
19	13.4	55.8	22	6	AX021010	AX021010 Sequence
20	13.4	55.8	24	6	AR066891	AR066891 Sequence
21	13.2	55.0	22	6	AR066609	AR066609 Sequence
22	13.2	55.0	23	6	I34867	I34867 Sequence 10
23	13	54.2	21	6	A26766	A26766 Oligonucleo
24	13	54.2	21	6	AR004454	AR004454 Sequence
25	12.8	53.3	17	6	AX475388	AX475388 Sequence
26	12.8	53.3	17	6	AX475390	AX475390 Sequence
27	12.8	53.3	20	12	AB068351	AB068351 Synthetic
28	12.8	53.3	21	6	AR066947	AR066947 Sequence
29	12.8	53.3	23	6	AX487008	AX487008 Sequence
30	12.8	53.3	24	6	AX164356	AX164356 Sequence
31	12.8	53.3	25	6	AX197195	AX197195 Sequence
32	12.8	53.3	25	6	AX476365	AX476365 Sequence
33	12.8	53.3	25	6	AX476375	AX476375 Sequence
34	12.6	52.5	22	6	AX033431	AX033431 Sequence
35	12.4	51.7	25	6	AR137526	AR137526 Sequence
36	12.4	51.7	25	6	AR139032	AR139032 Sequence
37	12.4	51.7	25	6	AX300826	AX300826 Sequence
38	12.2	50.8	20	6	A57378	A57378 Sequence 15
39	12.2	50.8	20	6	A61160	A61160 Sequence 1
40	12.2	50.8	20	6	A83993	A83993 Sequence 14
41	12.2	50.8	20	6	AR020539	AR020539 Sequence
42	12.2	50.8	20	6	AR089400	AR089400 Sequence
43	12.2	50.8	20	6	AR093600	AR093600 Sequence
44	12.2	50.8	20	6	AX030695	AX030695 Sequence
45	12.2	50.8	20	6	AX047253	AX047253 Sequence

ALIGNMENTS

RESULT 1  
AX117664  
LOCUS  
DEFINITION  
Sequence 2787 from Patent WO0129262.  
ACCESSION  
AX117664  
VERSION  
AX117664.1 GI:14034615  
KEYWORDS  
synthetic construct.  
artificial construct.  
artificial sequences.  
1 (bases 1 to 25)  
Picoult-Newburg, L. and Pohl, M.  
Genotyping reagents, kits and methods of use thereof  
Patent: WO 0129262-A 2787 26-APR-2001;  
Orchid Biosciences, Inc. (US)

AX117664  
Sequence 2787 from Patent WO0129262.  
AX117664  
AX117664.1 GI:14034615  
synthetic construct.  
artificial construct.  
artificial sequences.  
1 (bases 1 to 25)  
Picoult-Newburg, L. and Pohl, M.  
Genotyping reagents, kits and methods of use thereof  
Patent: WO 0129262-A 2787 26-APR-2001;  
Orchid Biosciences, Inc. (US)

Pred. No. is the number of results predicted by chance to have a

```
FEATURES
  source      Location/Qualifiers
  1..25
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="Primer"
BASE COUNT   10 a      8 c      6 g      1 t
ORIGIN

Query Match      66.7%; Score 16; DB 6; Length 25;
Best Local Similarity 79.2%; Pred. No. 5.9e+03;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAGAGCAGAAAGACACGCAAGCC 24
    ||||| ||||| || |||||
Db 1 GAGAGCAGAAAGACACGCAAGCC 24

RESULT 2
AX452262
LOCUS      21 bp      DNA      linear      PAT 06-JUL-2002
DEFINITION Sequence 19 from Patent WO0242442.
ACCESSION AX452262
VERSION   AX452262.1 GI:21712192
KEYWORDS  .
SOURCE    synthetic construct.
          synthetic construct
          artificial sequences.
ORGANISM  1
REFERENCE Baranyi,F., Braunschweiler,M., Gasch,A. and Berghof,K.
AUTHORS  Novel yeast strain for consumption
TITLE    Patent: WO 0242442-A 19 30-MAY-2002;
JOURNAL  Bioteccon Diagnostics GmbH (DE)
FEATURES  Location/Qualifiers
  source      1..21
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="Primer zur Differenzierung phylogenetischen
Einheiten, wie Stammen, Unterstammen, Spezies"
BASE COUNT   7 a      9 c      4 g      1 t
ORIGIN

Query Match      63.3%; Score 15.2; DB 6; Length 21;
Best Local Similarity 85.0%; Pred. No. 1.5e+04;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAAGAAACGACGCAAGCC 24
    ||||| || |||||
Db 2 GCAAGAAACGACGCAAGCC 21

RESULT 3
AX288411/c
LOCUS      24 bp      DNA      linear      PAT 21-NOV-2001
DEFINITION Sequence 173 from Patent WO0179548.
ACCESSION AX288411
VERSION   AX288411.1 GI:17050094
KEYWORDS  .
SOURCE    synthetic construct.
          synthetic construct
          artificial sequences.
ORGANISM  1
REFERENCE Baranyi,F., Zirvi,M., Gerry,N.P., Favis,R. and Kliman,R.
AUTHORS  Method of designing addressable array for detection of nucleic acid
TITLE    sequence differences using ligase detection reaction
JOURNAL  Patent: WO 0179548-A 173 25-OCT-2001;
          CORNELL RESEARCH FOUNDATION, INC. (US)
FEATURES  Location/Qualifiers
  source      1..24
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="Hypothetical Probe Sequence"
BASE COUNT   1 a      9 c      4 g      10 t
ORIGIN

Query Match      62.5%; Score 15; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GAAAGCACGCAAGCC 23
    ||||| |||||
Db 18 GAAAGCACGCAAGCC 4

RESULT 5
AX429796
LOCUS      25 bp      DNA      linear      PAT 21-JUN-2002
DEFINITION Sequence 24 from Patent EP1203826.
ACCESSION AX429796
VERSION   AX429796.1 GI:21540972
KEYWORDS  .
SOURCE    synthetic construct.
          synthetic construct
          artificial sequences.
ORGANISM  1
REFERENCE Ishizuka,T., Ishiguro,T. and Saitoh,J.
AUTHORS  Oligonucleotide for detection of hiv-1 and detection method
TITLE    Patent: EP 1203826-A 24 08-MAY-2002;
JOURNAL  Tosoh Corporation (JP)
FEATURES  Location/Qualifiers
  source      1..25
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="Primer"
BASE COUNT   12 a      6 c      6 g      1 t
ORIGIN

Query Match      60.0%; Score 14.4; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 3.6e+04;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 AGAAGCACGCAAGCC 23
    ||||| |||||
Db 1 AAAAGCACGCAAGCC 16
```

```

RESULT 6
LOCUS AX191307 20 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 4 from Patent WO0149880.
ACCESSION AX191307
VERSION AX191307.1 GI:15209558
KEYWORDS
SOURCE
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1587 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source
Location/Qualifiers
1..25
/db_xref="taxon:9606"
BASE COUNT 7 a 4 c 8 g 1 t
ORIGIN
Query Match 59.2%; Score 14.2; DB 6; Length 20;
Best Local Similarity 84.2%; Pred. No. 4.5e+04;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 GAGCAAGAGGACAGCAAA 21
Db 1 GAGCGGGAGGACAGCAAA 19
||||| ||| ||||| |||||

RESULT 7
LOCUS AX475389/c 17 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 610 from Patent WO0224750.
ACCESSION AX475389
VERSION AX475389.1 GI:22214674
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 610 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source
Location/Qualifiers
1..17
/db_xref="taxon:9606"
BASE COUNT 0 a 5 c 5 g 7 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 17;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
Db 17 AGAAAGCAGCAGCAGCC 1
||||| ||| ||||| |||||

RESULT 8
LOCUS AX476366/c 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1587 from Patent WO0224750.
ACCESSION AX476366
VERSION AX476366.1 GI:22215651
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1589 28-MAR-2002;
Aeomica, Inc. (US)

```

```

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1587 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source
Location/Qualifiers
1..25
/db_xref="taxon:9606"
BASE COUNT 0 a 10 c 5 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
Db 25 AGAAAGCAGCAGCAGCC 9
||||| ||| ||||| |||||

RESULT 9
LOCUS AX476367/c 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1588 from Patent WO0224750.
ACCESSION AX476367
VERSION AX476367.1 GI:22215652
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1588 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source
Location/Qualifiers
1..25
/db_xref="taxon:9606"
BASE COUNT 0 a 9 c 6 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
Db 24 AGAAAGCAGCAGCAGCC 8
||||| ||| ||||| |||||

RESULT 10
LOCUS AX476368/c 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1589 from Patent WO0224750.
ACCESSION AX476368
VERSION AX476368.1 GI:22215653
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1589 28-MAR-2002;
Aeomica, Inc. (US)

```

```

      Aeomica, Inc. (US)
FEATURES      Location/Qualifiers
  source      1..25
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   0 a 9 c 6 g 10 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   21 AGAAGCACAGCCAGCC 5

RESULT 13
AX476371/c
LOCUS      25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1592 from Patent WO0224750.
ACCESSION  AX476371
VERSION     AX476371.1 GI:22215656
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Zhang, J.
  TITLE     Human kidney tumor overexpressed membrane protein 1
  JOURNAL   Patent: WO 0224750-A 1592 28-MAR-2002;
            Aeomica, Inc. (US)
FEATURES
  source    1..25
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  2 a 8 c 6 g 9 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   20 AGAAGCACAGCCAGCC 4

RESULT 14
AX476372/c
LOCUS      25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1593 from Patent WO0224750.
ACCESSION  AX476372
VERSION     AX476372.1 GI:22215657
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Zhang, J.
  TITLE     Human kidney tumor overexpressed membrane protein 1
  JOURNAL   Patent: WO 0224750-A 1593 28-MAR-2002;
            Aeomica, Inc. (US)
FEATURES
  source    1..25
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  2 a 8 c 6 g 9 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   19 AGAAGCACAGCCAGCC 3

RESULT 15
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   19 AGAAGCACAGCCAGCC 3

      Aeomica, Inc. (US)
FEATURES      Location/Qualifiers
  source      1..25
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   0 a 9 c 6 g 10 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   23 AGAAGCACAGCCAGCC 7

RESULT 11
AX476369/c
LOCUS      25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1590 from Patent WO0224750.
ACCESSION  AX476369
VERSION     AX476369.1 GI:22215654
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Zhang, J.
  TITLE     Human kidney tumor overexpressed membrane protein 1
  JOURNAL   Patent: WO 0224750-A 1590 28-MAR-2002;
            Aeomica, Inc. (US)
FEATURES
  source    1..25
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  0 a 10 c 6 g 9 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db   22 AGAAGCACAGCCAGCC 6

RESULT 12
AX476370/c
LOCUS      25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1591 from Patent WO0224750.
ACCESSION  AX476370
VERSION     AX476370.1 GI:22215655
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Zhang, J.
  TITLE     Human kidney tumor overexpressed membrane protein 1
  JOURNAL   Patent: WO 0224750-A 1591 28-MAR-2002;
            Aeomica, Inc. (US)
FEATURES
  source    1..25
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  1 a 9 c 6 g 9 t
ORIGIN
Query Match  57.5%; Score 13.8; DB 6; Length 25;
```

AX476373/c  
LOCUS AX476373 25 bp DNA linear PAT 12-AUG-2002  
DEFINITION Sequence 1594 from Patent WO0224750.  
ACCESSION AX476373  
VERSION AX476373.1 GI:22215658  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Zhang, J.  
TITLE Human kidney tumor overexpressed membrane protein 1  
JOURNAL Patent: WO 0224750-A 1594 28-MAR-2002;  
Aecomica, Inc. (US)  
FEATURES  
source 1..25  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 2 a 8 c 6 g 9 t  
ORIGIN  
Query Match 57.5%; Score 13.8; DB 6; Length 25;  
Best Local Similarity 88.2%; Pred. No. 7e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 8 AGAAGCACGACGAGCC 24  
||||||| ||| ||||  
Db 18 AGAAGCACGAGCCAGCC 2

Search completed: July 17, 2003, 21:43:07  
Job time : 1862 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 19:54:17 ; Search time 227 Seconds  
(without alignments)  
236.097 Million cell updates/sec

Title: US-09-923-684-12  
Perfect score: 24  
Sequence: 1 gagagcaagaaagcacagcaagcc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 1698378

Minimum DB seq length: 0  
Maximum DB seq length: 25

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 25: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	24	Human SIM2 antisense
2	16	66.7	25	22	SNP specific SNPE
3	15.2	63.3	21	24	Yeast PCR primer S
4	15.2	63.3	24	24	Capture oligonucle
5	15.2	63.3	24	24	Capture oligonucle
6	15.2	63.3	24	24	Capture oligonucle
7	15.2	63.3	24	24	Capture oligonucle
8	15	62.5	20	24	Human BUB1A kinase
9	14.4	60.0	25	20	AAH58973

10 14.4 60.0 25 24 ABK85446  
11 14.2 59.2 20 22 AAD21350  
12 13.8 57.5 17 24 ABQ63834  
13 13.8 57.5 25 24 ABQ64811  
14 13.8 57.5 25 24 ABQ64812  
15 13.8 57.5 25 24 ABQ64813  
16 13.8 57.5 25 24 ABQ64814  
17 13.8 57.5 25 24 ABQ64815  
18 13.8 57.5 25 24 ABQ64816  
19 13.8 57.5 25 24 ABQ64817  
20 13.8 57.5 25 24 ABQ64818  
21 13.8 57.5 25 24 ABQ64819  
22 13.6 56.7 22 20 AAX00990  
23 13.4 55.8 22 20 AAX88559  
24 13.4 55.8 24 16 AAX82239  
25 13.2 55.0 22 18 AAV19325  
26 13.2 55.0 23 16 AAO84605  
27 13.2 55.0 23 24 ABA04462  
28 13.2 55.0 25 24 ABT03587  
29 13 54.2 18 21 AAZ71610  
30 13 54.2 21 14 AAQ43004  
31 13 54.2 25 24 AAL37715  
32 12.8 53.3 17 20 AAI83887  
33 12.8 53.3 17 24 ABQ63833  
34 12.8 53.3 17 24 ABQ63835  
35 12.8 53.3 20 24 AAL43431  
36 12.8 53.3 20 24 ABL44638  
37 12.8 53.3 21 16 AAQ82295  
38 12.8 53.3 23 22 AAH73895  
39 12.8 53.3 24 22 AAH55942  
40 12.8 53.3 25 22 AAI62271  
41 12.8 53.3 25 24 ABQ64820  
42 12.8 53.3 25 24 ABQ64810  
43 12.6 52.5 21 23 AAH89899  
44 12.6 52.5 22 21 AAA73720  
45 12.6 52.5 22 24 ABL51638

# ALIGNMENTS

RESULT 1  
AAD30531  
ID AAD30531 standard; DNA; 24 BP.  
XX  
AC AAD30531;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Human SIM2 antisense oligonucleotide #2.  
XX  
KW Human; single minded homologue 2; SIM2; colon cancer; prostate cancer;  
XX  
KW pancreas cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200212565-A1.  
XX  
PD 14-FEB-2002.  
XX  
PF 06-AUG-2001; 2001WO-US24781.  
XX  
PR 04-AUG-2000; 2000US-223531P.  
PR 22-DEC-2000; 2000US-257965P.  
XX  
XX (UYFL ) UNIV FLORIDA ATLANTIC.  
XX  
XX Narayanan R;  
XX  
XX WPI; 2002-227169/28.  
XX  
XX Detecting cancer in tissue sample, involves providing tissue sample,  
XX and analyzing tissue sample for presence of single minded homologue 2

HIV-1 RNA, second  
NASBA reaction ass  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Human KTM1a portl  
Mus musculus notch  
Human chromosome 1  
Chromosome 11 (loc  
Type 3 sense PCR I  
D-loop mtDNA prime  
Human P1030 PCR p  
Human En-1 gene PC  
Human biallelic ma  
PCR primer PIA-1 t  
N-Acetylglucosamin  
Human TIE-2 substr  
Human KTM1a portl  
Human KTM1a portl  
Carica papaya gend  
Human chromosome 1  
Chromosome 11 (loc  
Human hepatome cel  
Human SCN1A PCR-SS  
Soybean 318013 reg  
Human KTM1a portl  
Human KTM1a portl  
Human polymorphic  
5' primer used to  
Human mitotic PITS

PT marker -  
 PS Claim 30; Page 58; 60pp; English.  
 CC The invention relates to a method of detecting cancer in a tissue  
 CC sample. The method involves providing the tissue sample and analysing  
 CC the tissue sample for the presence of a single minded homologue 2 (SIM2)  
 CC nucleic acid or protein, where the presence of the SIM2 marker in the  
 CC tissue sample indicates that the tissue sample contains cancer. The  
 CC method is useful for identifying compounds that modulate expression of  
 CC SIM2 gene in a cell. It is useful for detecting and reducing the growth  
 CC of cancer such as colon, prostate and pancreas cancer in an animal or  
 CC mammal. The present sequence is human SIM2 antisense oligonucleotide  
 CC used in the exemplification of the invention.  
 XX Sequence 24 BP; 11 A; 6 C; 7 G; 0 U; 0 other;  
 SQ  
 Query Match 100.0%; Score 24; DB 24; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 0.68; 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GAGACGAAGAAGCAGCAGCAGCC 24  
 Db 1 GAGACGAAGAAGCAGCAGCAGCC 24  
 RESULT 2  
 AAH39991  
 ID AAH39991 standard; DNA; 25 BP.  
 XX AC AAH39991;  
 XX 14-AUG-2001 (first entry)  
 XX SNP specific SNPE primer SEQ ID 2787.  
 XX Single nucleotide polymorphism: SNP; single nucleotide primer extension;  
 KW SNPE; genotyping; agammaglobulinaemia; diabetes insipidus; cancer;  
 KW Lesch-Nyhan syndrome; muscular dystrophy; familial hypercholesterolaemia;  
 KW polycystic kidney disease; osteogenesis imperfecta; autoimmune disease;  
 KW acute intermittent porphyria; rheumatoid arthritis; multiple sclerosis;  
 KW inflammation; forensic investigation; paternity analysis; primer; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200129262-A2.  
 XX 26-APR-2001.  
 XX 13-OCT-2000; 2000WO-US28436.  
 XX 15-OCT-1999; 99US-0160096.  
 XX (ORCH-) ORCHID BIOSCIENCES INC.  
 XX Picoult-Newburg L, Pohl M;  
 XX WPI; 2001-290930/30.  
 XX New genotyping oligonucleotide, useful for detecting the presence,  
 PT absence or identity of single polynucleotide polymorphism in a nucleic  
 PT acid sample -  
 XX Claim 1; Page 64; 83pp; English.  
 PS Sequences AAH37205 - AAH40944 represent PCR primers, single nucleotide  
 CC primer extension (SNPE) primers, and the sequences of regions flanking  
 CC sites of single nucleotide polymorphisms SNPs. The present invention  
 CC includes kits for determining the presence or absence of a SNP, using the  
 CC oligonucleotides of the invention. The PCR primers are used to amplify a  
 CC SNP flanking sequence, the SNPE primer is used as a genotyping primer.  
 CC The oligonucleotides are useful for genotyping a nucleic acid sample by  
 CC performing a single-nucleotide primer extension reaction. The

CC oligonucleotides are useful for determining the presence, absence or  
 CC identity of a SNP and for genotyping nucleic acid samples, for e.g. to  
 CC assess by association analysis the genotype of an individual or group of  
 CC individuals, having a pathological phenotypic trait suspected of being  
 CC caused by one or more SNPs. Phenotypic traits include diseases e.g.  
 CC agammaglobulinaemia, diabetes insipidus, Lesch-Nyhan syndrome, muscular  
 CC dystrophy, familial hypercholesterolaemia, polycystic kidney disease,  
 CC osteogenesis imperfecta and acute intermittent porphyria. Phenotypic  
 CC traits also include symptoms of or susceptibility to multifactorial  
 CC disease of which a component is or may be genetic such as autoimmune  
 CC diseases, including, rheumatoid arthritis, multiple sclerosis,  
 CC inflammation, cancer, nervous system diseases and infection by pathogenic  
 CC microorganism. The method is also useful in forensic investigations and  
 CC paternity analysis. The present sequence represents a single nucleotide  
 CC primer extension (SNPE) primer specific for a human SNP containing DNA  
 CC sequence.  
 XX Sequence 25 BP; 10 A; 8 C; 6 G; 1 T; 0 other;  
 SQ  
 Query Match 66.7%; Score 16; DB 22; Length 25;  
 Best Local Similarity 79.2%; Pred. No. 1.4e+03;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 GAGACGAAGAAGCAGCAGCAGCC 24  
 Db 1 GAGACGAAGAAGCAGCAGCAGCC 24  
 RESULT 3  
 ABN81518  
 ID ABN81518 standard; DNA; 21 BP.  
 XX AC ABN81518;  
 XX 13-AUG-2002 (first entry)  
 XX Yeast PCR primer SEQ ID NO 19.  
 XX Yeast; pharmaceutical; diarrhoea; intestinal infection; Candida;  
 KW fermented drink; antidiarrhoeic; fungicide; antibacterial;  
 KW dermatological; gastrointestinal; PCR; primer; ss.  
 XX Synthetic.  
 OS  
 XX WO200242442-A2.  
 XX 30-MAY-2002.  
 XX 15-OCT-2001; 2001WO-EP11887.  
 XX 24-NOV-2000; 2000DE-1058379.  
 XX (BIOT-) BIOTECON DIAGNOSTICS GMBH.  
 XX Grabowski R, Braunschweiler M, Gasch A, Berghof K;  
 XX WPI; 2002-463630/49.  
 XX New yeast strains characterized by specific band patterns in a  
 PT polymerase chain reaction, useful e.g. as probiotics or for preparing  
 PT fermented drinks -  
 XX Claim 2; Page 8; 29pp; German.  
 PS The invention relates to yeast strains (A) that produce a specific band  
 CC pattern, illustrated in the specification, when characterised by a  
 CC polymerase chain reaction (PCR). (A), optionally in lyophilised form or  
 CC as extracts or culture supernatants, are useful for administration to  
 CC humans or animals, as pharmaceuticals (for treating diarrhoea, colitis,  
 CC intestinal infections, Candida infections, or skin disorders) or  
 CC probiotics, also for preparation of fermented drinks, suspensions,  
 CC extracts and baked goods. (A) have only a minimal effect on the taste of  
 CC goods prepared using them and can be unequivocally identified by

CC genetic characterisation, even though they are nearly impossible to  
 CC differentiate biochemically. The present sequence is that of a PCR primer  
 CC of the invention.

XX Sequence 21 BP; 7 A; 9 C; 4 G; 1 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 21;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCAGAAAGCACAGCAAGCC 24  
 ||||| ||| |||||  
 Db 2 GCAGAAATCCAGCAAGCC 21

## RESULT 4

ABI82742/C  
 ID ABI82742 standard; DNA; 24 BP.

XX ABI82742;

DT 15-FEB-2002 (first entry)

XX Capture oligonucleotide Zip ID#173 oligo #1.

Human; K-ras; PCR primer; probe; capture probe; mutation detection;  
 ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;  
 infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;  
 cancer; oncogene; tumour suppressor; human papillomavirus; forensic;  
 environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX WO200179548-A2.

XX 25-OCT-2001.

XX 04-APR-2001; 2001WO-US10958.

XX 14-APR-2000; 2000US-197271P.

XX (CORR ) CORNELL RES FOUND INC.

XX Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX WPI; 2002-034366/04.

Designing capture oligonucleotide probes for use on a support to which  
 complementary oligonucleotides hybridize with little mismatch -

Example 5; Fig 25; 300pp; English.

The present invention describes a method (M1) for designing capture  
 oligonucleotide probes (I) for use on a support to which complementary  
 oligonucleotide probes (II) will hybridize with little mismatch, where  
 (I) have melting temperatures within a narrow range. The method is useful  
 for detecting infectious diseases caused by bacterial infectious agents  
 e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal  
 infectious agents e.g. Cryptococcus neoformans, Candida albicans and  
 Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,  
 Epstein-Barr virus and polio virus, and parasitic infectious agents  
 selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus  
 as 21 hydroxylase deficiency. The method is also useful for detecting genetic diseases such  
 as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.

Involving in DNA amplification, replication, recombination or repair, the  
 cancer is specifically associated with a gene selected from BRCA1 gene,  
 p53 gene, human papillomavirus types 16 and 18 and liver cancers. The  
 method is also used for environmental monitoring, forensics and the food  
 and feed industry, detecting comprises scanning (using e.g. a scanning  
 electron microscope and infrared microscope) the support at the  
 particular sites and identifying if ligation of the oligonucleotide probe  
 sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to  
 CC ABI97546 represent oligonucleotide sequences used in the exemplification  
 CC of the present invention.

XX Sequence 24 BP; 1 A; 9 C; 4 G; 10 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 24;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 AGCAAGAAAGCACAGCAAGC 23  
 || ||||| |||||  
 Db 23 AGGGGAAAGCACAGCAAGC 4

## RESULT 5

ABI82743  
 ID ABI82743 standard; DNA; 24 BP.

XX ABI82743;

DT 15-FEB-2002 (first entry)

XX Capture oligonucleotide Zip ID#173 oligo #2.

Human; K-ras; PCR primer; probe; capture probe; mutation detection;  
 ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;  
 infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;  
 cancer; oncogene; tumour suppressor; human papillomavirus; forensic;  
 environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX WO200179548-A2.

XX 25-OCT-2001.

XX 04-APR-2001; 2001WO-US10958.

XX 14-APR-2000; 2000US-197271P.

XX (CORR ) CORNELL RES FOUND INC.

XX Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX WPI; 2002-034366/04.

Designing capture oligonucleotide probes for use on a support to which  
 complementary oligonucleotides hybridize with little mismatch -

Example 5; Fig 25; 300pp; English.

The present invention describes a method (M1) for designing capture  
 oligonucleotide probes (I) for use on a support to which complementary  
 oligonucleotide probes (II) will hybridize with little mismatch, where  
 (I) have melting temperatures within a narrow range. The method is useful  
 for detecting infectious diseases caused by bacterial infectious agents  
 e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal  
 infectious agents e.g. Cryptococcus neoformans, Candida albicans and  
 Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,  
 Epstein-Barr virus and polio virus, and parasitic infectious agents  
 selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus  
 as 21 hydroxylase deficiency. The method is also useful for detecting genetic diseases such  
 as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.

Involving in DNA amplification, replication, recombination or repair, the  
 cancer is specifically associated with a gene selected from BRCA1 gene,  
 p53 gene, human papillomavirus types 16 and 18 and liver cancers. The  
 method is also used for environmental monitoring, forensics and the food  
 and feed industry, detecting comprises scanning (using e.g. a scanning  
 electron microscope and infrared microscope) the support at the  
 particular sites and identifying if ligation of the oligonucleotide probe  
 sets occurred and correlating (using a computer) identified ligation to a

Mon Jul 21 15:46:53 2003

us-09-923-684-12.szlm25.rng

CC presence or absence of the target nucleotide sequences. ABI82074 to  
 CC ABI97546 represent oligonucleotide sequences used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 24 BP; 10 A; 4 C; 9 G; 1 T; 0 other;  
 Query Match 63.3%; Score 15.2; DB 24; Length 24;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 AGCAAGAAAGCAGCAAGC 23  
 || |||||  
 Db 2 AGGGGAAAGCAGCAAGC 21

RESULT 6  
 ABI92008/C  
 ID ABI92008 standard; DNA; 24 BP.  
 XX  
 AC ABI92008;  
 XX  
 DT 15-FEB-2002 (first entry)  
 XX  
 DE Capture oligonucleotide Zip ID#173 oligo #3.  
 XX  
 KW Human; K-ras; PCR primer; probe; capture probe; mutation detection;  
 KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;  
 KW infection; 21 hydroxylase deficiency; Turner Syndrome; Obesity;  
 KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;  
 KW environmental monitoring; food industry; feed industry; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200179548-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 04-APR-2001; 2001WO-US10958.  
 XX  
 PR 14-APR-2000; 2000US-197271P.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;  
 XX  
 WPI; 2002-034366/04.  
 XX  
 DR Designing capture oligonucleotide probes for use on a support to which  
 XX complementary oligonucleotides hybridize with little mismatch -  
 PT  
 PS Claim 3; Fig 26; 300pp; English.  
 XX  
 CC The present invention describes a method (M1) for designing capture  
 CC oligonucleotide probes (I) for use on a support to which complementary  
 CC oligonucleotide probes (II) will hybridize with little mismatch, where  
 CC (I) have melting temperatures within a narrow range. The method is useful  
 CC for detecting infectious diseases caused by bacterial infectious agents  
 CC e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal  
 CC infectious agents e.g. Cryptococcus neoformans, Candida albicans and  
 CC Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,  
 CC Epstein-Barr virus and polio virus, and parasitic infectious agents  
 CC selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus  
 CC medinis. The method is also useful for detecting genetic diseases such  
 CC as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.  
 CC Detecting cancer involving oncogenes, tumour suppressor genes, or genes  
 CC involved in DNA amplification, replication, recombination or repair, the  
 CC cancer is specifically associated with a gene selected from BRCA1 gene,  
 CC p53 gene, human papillomavirus types 16 and 18 and liver cancers. The  
 CC method is also used for environmental monitoring, forensics and the food  
 CC and feed industry, detecting comprises scanning (using e.g. a scanning  
 CC electron microscope and infrared microscope) the support at the  
 CC particular sites and identifying if ligation of the oligonucleotide probe  
 CC sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to  
 CC ABI97546 represent oligonucleotide sequences used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 24 BP; 1 A; 9 C; 4 G; 10 T; 0 other;  
 Query Match 63.3%; Score 15.2; DB 24; Length 24;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 AGCAAGAAAGCAGCAAGC 23  
 || |||||  
 Db 23 AGGGGAAAGCAGCAAGC 4

RESULT 7  
 ABI92009  
 ID ABI92009 standard; DNA; 24 BP.  
 XX  
 AC ABI92009;  
 XX  
 DT 15-FEB-2002 (first entry)  
 XX  
 DE Capture oligonucleotide Zip ID#173 oligo #4.  
 XX  
 KW Human; K-ras; PCR primer; probe; capture probe; mutation detection;  
 KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;  
 KW infection; 21 hydroxylase deficiency; Turner Syndrome; Obesity;  
 KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;  
 KW environmental monitoring; food industry; feed industry; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200179548-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 04-APR-2001; 2001WO-US10958.  
 XX  
 PR 14-APR-2000; 2000US-197271P.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;  
 XX  
 WPI; 2002-034366/04.  
 XX  
 DR Designing capture oligonucleotide probes for use on a support to which  
 XX complementary oligonucleotides hybridize with little mismatch -  
 PT  
 PS Claim 3; Fig 26; 300pp; English.  
 XX  
 CC The present invention describes a method (M1) for designing capture  
 CC oligonucleotide probes (I) for use on a support to which complementary  
 CC oligonucleotide probes (II) will hybridize with little mismatch, where  
 CC (I) have melting temperatures within a narrow range. The method is useful  
 CC for detecting infectious diseases caused by bacterial infectious agents  
 CC e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal  
 CC infectious agents e.g. Cryptococcus neoformans, Candida albicans and  
 CC Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,  
 CC Epstein-Barr virus and polio virus, and parasitic infectious agents  
 CC selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus  
 CC medinis. The method is also useful for detecting genetic diseases such  
 CC as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.  
 CC Detecting cancer involving oncogenes, tumour suppressor genes, or genes  
 CC involved in DNA amplification, replication, recombination or repair, the  
 CC cancer is specifically associated with a gene selected from BRCA1 gene,  
 CC p53 gene, human papillomavirus types 16 and 18 and liver cancers. The  
 CC method is also used for environmental monitoring, forensics and the food  
 CC and feed industry, detecting comprises scanning (using e.g. a scanning  
 CC electron microscope and infrared microscope) the support at the  
 CC particular sites and identifying if ligation of the oligonucleotide probe  
 CC sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to  
CC ABI97546 represent oligonucleotide sequences used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 24 BP; 10 A; 4 C; 9 G; 1 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 24;  
Best Local Similarity 85.0%; Pred. No. 2.9e+03;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AGCAAGAAAGACACACAGC 23  
|| |||||  
Db 2 AGGGGGAAGACACACAGC 21

## RESULT 8

ABI93086/C  
ID ABI93086 standard; DNA; 20 BP.

AC ABI93086;

DT 15-FEB-2002 (first entry)

DE Capture oligonucleotide Zip ID#173 oligo #9.

KW Human: K-ras; PCR primer; probe; capture probe; mutation detection;  
KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;  
KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;  
KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;  
KW environmental monitoring; food industry; feed industry; ss.

OS Synthetic.

PN WO200179548-A2.

XX 25-OCT-2001.

PF 04-APR-2001; 2001WO-US10958.

PR 14-APR-2000; 2000US-197271P.

PA (CORR ) CORNELL RES FOUND INC.

PI Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

DR WPI; 2002-034366/04.

PT Designing capture oligonucleotide probes for use on a support to which  
PT complementary oligonucleotides hybridize with little mismatch -

PS Example 5; Fig 29; 300pp; English.

CC The present invention describes a method (M1) for designing capture  
CC oligonucleotide probes (I) for use on a support to which complementary  
CC oligonucleotide probes (II) will hybridize with little mismatch, where  
CC (I) have melting temperatures within a narrow range. The method is useful  
CC for detecting infectious diseases caused by bacterial infectious agents  
CC e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal  
CC infectious agents e.g. Cryptococcus neoformans, Candida albicans and  
CC Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,  
CC Epstein-Barr virus and polio virus, and parasitic infectious agents  
CC selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus  
CC medinensis. The method is also useful for detecting genetic diseases such  
CC as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.  
CC Detecting cancer involving oncogenes, tumour suppressor genes, or genes  
CC involved in DNA amplification, replication, recombination or repair, the  
CC cancer is specifically associated with a gene selected from BRCA1 gene,  
CC p53 gene, human papillomavirus types 16 and 18 and liver cancers. The  
CC method is also used for environmental monitoring, forensics and the food  
CC and feed industry, detecting comprises scanning (using e.g. a scanning  
CC electron microscope and infrared microscope) the support at the  
CC particular sites and identifying if ligation of the oligonucleotide probe  
CC sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to  
CC ABI97546 represent oligonucleotide sequences used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 20 BP; 0 A; 7 C; 4 G; 9 T; 0 other;

Query Match 62.5%; Score 15; DB 24; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.5e+03;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GAAGACACACAGC 23  
|||||  
Db 18 GAAGACACACAGC 4

## RESULT 9

AAAX58973/C

ID AAAX58973 standard; DNA; 25 BP.

XX AAAX58973;

AC AAAX58973;

DT 23-AUG-1999 (first entry)

DE Human BUB1A kinase 5'RACE primer TJY 300.

KW BUB1A; human; protein kinase; checkpoint control; mitosis;  
KW kinetochore; cell proliferation; cancer; tumour; therapy; PCR;

KW RACE; primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9928334-A1.

XX 10-JUN-1999.

PF 01-DEC-1998; 98WO-US25415.

PR 01-DEC-1997; 97US-0067093.

PA (FOXC-) FOX CHASE CANCER CENT.

PI Chan G, Jablonski S, Yen T;

DR WPI; 1999-371094/31.

PT Human BUB genes and proteins involved in mitotic checkpoint control

PS Disclosure; Page 31; 99pp; English.

CC This primer oligonucleotide, termed TJY 300, was used with nested  
CC 5' Clontech primers AP1 and AP2 to isolate full-length human BUB1A  
CC cDNA (see AAAX58970) by extending the yeast two-hybrid Int-57 clone  
CC towards the 5' end by RACE. Human BUB1A (see AAY06286) is a novel  
CC kinetochore kinase that is involved in mitotic checkpoint control.  
CC BUB genes (see AAAX58970-72) and their encoded proteins (see AAY06286-88)  
CC disclosed in the invention provide valuable therapeutic targets for  
CC the design of antiproliferative agents which inhibit the aberrant  
CC cellular proliferation observed in tumour cells. Antibodies and  
CC oligonucleotide probes based on BUB are also useful in screening  
CC for therapeutics, diagnosis and identification of homologues.

XX Sequence 25 BP; 2 A; 8 C; 4 G; 11 T; 0 other;

Query Match 60.0%; Score 14.4; DB 20; Length 25;  
Best Local Similarity 93.8%; Pred. No. 6.3e+03;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GAGCAAGAAAGACAG 18  
|||||  
Db 25 GAGCAAGAAAGACAG 10

```
RESULT 10
ABK85446
ID ABR85446 standard; DNA; 25 BP.
XX
AC ABR85446;
XX
DT 14-AUG-2002 (first entry)
XX
DE HIV-1 RNA, second primer #6.
XX
KW Human immunodeficiency virus type 1; HIV-1 detection method;
KW primer; ss.
XX
OS Human immunodeficiency virus type 1.
XX
PN EP1203826-A2.
XX
PD 08-MAY-2002.
XX
PF 30-OCT-2001; 2001EP-0125378.
XX
PR 30-OCT-2000; 2000JP-0334937.
XX
PA (TOYU) TOSOH CORP.
XX
PI Ishizuka T, Ishiguro T, Saitoh J;
XX
PWPI; 2002-473032/51.
XX
XX An oligonucleotide useful for detection of an RNA derived from HIV-1 in
XX clinical tests and diagnosis.
XX
XX Claim 4; Page 21; 34pp; English.
XX
XX The present invention relates to oligonucleotides binding to specific
XX sites of human immunodeficiency virus type 1 (HIV-1) RNA. The
XX oligonucleotides are useful for detecting HIV-1 in clinical tests
XX and diagnosis. The oligonucleotides provide simple, speedy and
XX sensitive detection of HIV-1 RNA which can bind to an intramolecularly
XX free region of the genomic RNA of HIV-1 at relatively low and constant
XX temperatures. The detection method comprises synthesizing a cDNA by
XX the action of an RNA-dependent DNA polymerase by using a specific
XX sequence in an RNA derived from HIV-1 anticipated in a sample as a
XX template, a first primer containing a sequence complementary to the
XX specific sequence and a second primer containing a sequence homologous
XX to the specific sequence (either of which additionally has a promoter
XX sequence for the RNA polymerase at the 5' end). ABR85441-ABK85449
XX represent second primers for HIV-1 RNA.
XX
XX Sequence 25 BP; 12 A; 6 C; 6 G; 1 T; 0 other;
XX
XX Query Match 60.0%; Score 14.4; DB 24; Length 25;
XX Best Local Similarity 93.8%; Pred. No. 6.3e+03;
XX Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 8 AGAAGCACAGCAAGC 23
XX 1 AAAAAGCACAGCAAGC 16
XX
XX Db
XX
XX RESULT 11
AAH21350
ID AAH21350 standard; DNA; 20 BP.
XX
AC AAH21350;
XX
DT 17-SEP-2001 (first entry)
XX
DE NASBA reaction associated primer SEQ ID 4.
XX
KW Primer; primer-dependent nucleic acid synthesis; amplification;
KW sequencing; cDNA synthesis; primer-based mutagenesis; ss.
XX
XX
```

```
Unidentified.
OS
XX WO200149880-A2.
XX
XX 12-JUL-2001.
XX
XX 27-DEC-2000; 2000WO-EP13288.
XX
XX 30-DEC-1999; 99DE-1063857.
XX
XX (QIAG-) QIAGEN GMBH.
XX
XX Korfhage C, Oelmueller U;
XX
XX WPI; 2001-441896/47.
XX
XX New primers, useful for nucleic acid extension and amplification
XX reactions, with specified base contents in different regions, have
XX increased selectivity and efficiency.
XX
XX Example 1; Page 28; 50pp; German.
XX
XX This invention describes a novel primer (I) in which the proportion of
XX bases that form two hydrogen bonds with the complementary target is at
XX least 50% within the last 6 bases before the 3'-end of the hybridizing
XX portion of (I) and the proportion of bases that form, similarly, three
XX hydrogen bonds is at least 60% within the last 1/4 from the 5'-end of
XX the hybridizing portion. (I), or their mixtures, are used for
XX primer-dependent nucleic acid synthesis and amplification, particularly
XX under isothermal conditions. Typical applications are in sequencing,
XX cDNA synthesis and primer-based mutagenesis. Compared with conventional
XX primers, (I) hybridize very specifically and efficiently to targets,
XX resulting in significantly greater yields with reduced formation of
XX by-products, and eliminate the need to compensate for poor specificity
XX by increasing the temperature. This sequence represents a NASBA nucleic
XX acid sequence based amplification primer used to illustrate the method
XX of the invention.
XX
XX Sequence 20 BP; 7 A; 4 C; 8 G; 1 T; 0 other;
XX
XX Query Match 59.2%; Score 14.2; DB 22; Length 20;
XX Best Local Similarity 84.2%; Pred. No. 7.5e+03;
XX Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 3 GAGCAAGAAAGCACAGCA 21
XX 1 GAGCGGGAAGGCACAGCAA 19
XX
XX Db
XX
XX RESULT 12
ABQ63834/C
ID ABQ63834 standard; DNA; 17 BP.
XX
XX AC ABQ63834;
XX
XX 20-AUG-2002 (first entry)
XX
XX DE Human KTOM1a portion (ABQ63232) probe # 547.
XX
XX KW Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200224750-A2.
XX
XX PD 28-MAR-2002.
XX
XX 21-SEP-2001; 2001WO-US29656.
XX
XX 21-SEP-2000; 2000US-234687P.
XX
XX 27-SEP-2000; 2000US-236359P.
XX
```

```

PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US000661.
PR 30-JAN-2001; 2001WO-US000662.
PR 30-JAN-2001; 2001WO-US000663.
PR 30-JAN-2001; 2001WO-US000664.
PR 30-JAN-2001; 2001WO-US000665.
PR 30-JAN-2001; 2001WO-US000666.
PR 30-JAN-2001; 2001WO-US000667.
PR 30-JAN-2001; 2001WO-US000668.
PR 30-JAN-2001; 2001WO-US000669.
PR 30-JAN-2001; 2001WO-US000670.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX
PA (AEOM-) AEOMICA INC.
XX
XX Zhang J;
XX
XX WPI; 2002-479509/51.
XX
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
XX nucleic acids encoding the protein, useful for treating subjects having
XX defects in KTOM1 which can manifest as cancer of the kidney, or as a
XX disorder of e.g., liver or bone
XX
XX Example 2; Page 229; 418pp; English.
XX
XX The invention relates to a novel isolated nucleic acid encoding human
XX KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
XX invention has cytostatic activity. The nucleotide may have a use in gene
XX therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
XX monitor a disease caused by altered expression of human KTOM1.
XX Compositions comprising the nucleic acids, proteins or antibodies may be
XX used to treat subjects having defects in KTOM1 which can manifest as
XX cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
XX heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
XX function. The sequence represents a probe used in the invention to
XX scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX
XX Sequence 17 BP; 0 A; 5 C; 5 G; 7 T; 0 other;
XX
XX Query Match 57.5%; Score 13.8; DB 24; Length 17;
XX Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
XX Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 8 AGAAAGCAGCAGCAGCC 24
XX 17 AGAAAGCAGCAGCAGCC 1
XX
XX Db
XX
XX RESULT 13
XX ABQ64811/c
XX ID ABQ64811 standard; DNA; 25 BP.
XX
XX AC ABQ64811;
XX
XX DT 20-AUG-2002 (first entry)
XX
XX DE Human KTOM1a portion (ABQ63232) probe # 1524.
XX
XX KW Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX KW kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200224750-A2.
XX
XX PD 28-MAR-2002.
XX
XX PF 21-SEP-2001; 2001WO-US29656.
XX
XX PR 21-SEP-2000; 2000US-234687P.
XX

```

---

```

PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US000661.
PR 30-JAN-2001; 2001WO-US000662.
PR 30-JAN-2001; 2001WO-US000663.
PR 30-JAN-2001; 2001WO-US000664.
PR 30-JAN-2001; 2001WO-US000665.
PR 30-JAN-2001; 2001WO-US000666.
PR 30-JAN-2001; 2001WO-US000667.
PR 30-JAN-2001; 2001WO-US000668.
PR 30-JAN-2001; 2001WO-US000669.
PR 30-JAN-2001; 2001WO-US000670.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX
XX (AEOM-) AEOMICA INC.
XX
XX Zhang J;
XX
XX WPI; 2002-479509/51.
XX
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
XX nucleic acids encoding the protein, useful for treating subjects having
XX defects in KTOM1 which can manifest as cancer of the kidney, or as a
XX disorder of e.g., liver or bone
XX
XX Example 2; Page 357; 418pp; English.
XX
XX The invention relates to a novel isolated nucleic acid encoding human
XX KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
XX invention has cytostatic activity. The nucleotide may have a use in gene
XX therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
XX monitor a disease caused by altered expression of human KTOM1.
XX Compositions comprising the nucleic acids, proteins or antibodies may be
XX used to treat subjects having defects in KTOM1 which can manifest as
XX cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
XX heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
XX function. The sequence represents a probe used in the invention to
XX scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX
XX Sequence 25 BP; 0 A; 10 C; 5 G; 10 T; 0 other;
XX
XX Query Match 57.5%; Score 13.8; DB 24; Length 25;
XX Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
XX Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 8 AGAAAGCAGCAGCAGCC 24
XX 25 AGAAAGCAGCAGCAGCC 9
XX
XX Db
XX
XX RESULT 14
XX ABQ64812/c
XX ID ABQ64812 standard; DNA; 25 BP.
XX
XX AC ABQ64812;
XX
XX DT 20-AUG-2002 (first entry)
XX
XX DE Human KTOM1a portion (ABQ63232) probe # 1525.
XX
XX KW Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX KW kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200224750-A2.
XX
XX PD 28-MAR-2002.
XX
XX PF 21-SEP-2001; 2001WO-US29656.
XX
XX PR 21-SEP-2000; 2000US-234687P.
XX

```

```

PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX (AEOM-) AEOMICA INC.
XX Zhang J;
XX WPI; 2002-479509/51.
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
XX nucleic acids encoding the protein, useful for treating subjects having
XX defects in KTOM1 which can manifest as cancer of the kidney, or as a
XX disorder of e.g., liver or bone
XX
XX Example 2; Page 357; 418pp; English.
XX The invention relates to a novel isolated nucleic acid encoding human
XX KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
XX invention has cytostatic activity. The nucleotide may have a use in gene
XX therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
XX monitor a disease caused by altered expression of human KTOM1.
XX Compositions comprising the nucleic acids, proteins or antibodies may be
XX used to treat subjects having defects in KTOM1 which can manifest as
XX cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
XX heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
XX function. The sequence represents a probe used in the invention to
XX scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX
XX Sequence 25 BP; 0 A; 9 C; 6 G; 10 T; 0 other;
XX
Query Match 57.5%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8 AGAAAGCAGCAGCAGCC 24
DB 24 AGAAAGCAGCAGCAGCC 8
RESULT 15
ABQ64813/c
ID ABQ64813 standard; DNA; 25 BP.
XX AC ABQ64813;
XX AC
XX 20-AUG-2002 (first entry)
XX Human KTOM1a portion (ABQ63232) probe # 1526.
XX Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX Homo sapiens.
XX WO200224750-A2.
XX 28-MAR-2002.
XX 21-SEP-2001; 2001WO-US29656.
XX PF

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:12:03 ; Search time 1638 Seconds  
(without alignments)  
237.297 Million cell updates/sec

Title: US-09-923-684-12  
Perfect score: 24  
Sequence: .1 gagagcaagaagcacagcaagcc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 18144

Minimum DB seq length: 0  
Maximum DB seq length: 25

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_estic: *
9: gb_est1: *
10: gb_est2: *
11: gb_est3: *
12: gb_est4: *
13: gb_est5: *
14: gb_est6: *
15: em_estfun: *
16: em_estcom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12.2	50.8	23	17	AZ469557
C 2	11.6	48.3	23	17	AZ434508
C 3	11.4	47.5	22	17	AZ937666
C 4	11.4	47.5	22	17	TA372612P
C 5	11.4	47.5	23	17	AZ486363
C 6	11.4	47.5	25	17	AZ648037

C 7 11.2 46.7 20 17 AZ453505  
C 8 11.2 46.7 23 17 AZ357282  
C 9 11.2 46.7 24 17 AZ999558  
C 10 11 45.8 19 17 AZ774205  
C 11 11 45.8 20 17 AZ834080  
C 12 11 45.8 24 17 AZ778465  
C 13 10.8 45.0 20 17 AZ796553  
C 14 10.8 45.0 20 17 AZ808800  
C 15 10.6 44.2 20 17 AZ308814  
C 16 10.6 44.2 22 9 AI808665  
C 17 10.6 44.2 22 17 AZ443174  
C 18 10.6 44.2 23 17 AZ318263  
C 19 10.4 43.3 21 17 AZ789335  
C 20 10.4 43.3 24 17 AZ402832  
C 21 10.4 43.3 24 17 AZ588136  
C 22 10.4 43.3 24 17 AZ815479  
C 23 10.2 42.5 21 17 AZ656028  
C 24 10.2 42.5 22 17 AZ941882  
C 25 10.2 42.5 23 17 AZ973926  
C 26 10.2 42.5 24 17 AZ307138  
C 27 10.2 42.5 24 17 AZ792444  
C 28 10.2 42.5 25 9 AI444500  
C 29 10.2 42.5 25 9 AA469268  
C 30 10.2 42.5 25 17 AZ389918  
C 31 10 41.7 21 17 AZ334510  
C 32 10 41.7 21 17 AZ336281  
C 33 10 41.7 22 17 AZ471500  
C 34 10 41.7 22 17 AZ466178  
C 35 10 41.7 25 9 AA988825  
C 36 10 41.7 25 17 AZ465511  
C 37 10 41.7 25 17 AZ810763  
C 38 9.8 40.8 19 17 AZ406137  
C 39 9.8 40.8 21 17 AZ308773  
C 40 9.8 40.8 21 17 AZ346717  
C 41 9.8 40.8 22 9 AI344786  
C 42 9.8 40.8 22 17 AZ307823  
C 43 9.8 40.8 22 17 AZ324102  
C 44 9.8 40.8 22 17 AZ950407  
C 45 9.8 40.8 23 17 AZ430288

## ALIGNMENTS

RESULT 1  
AZ469557/c

LOCUS  
DEFINITION

ACCESSION  
VERSION

KEYWORDS  
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ469557 23 bp DNA linear GSS 04-OCT-2000  
M0283A09F Mouse 10kb plasmid UUGCIM library Mus musculus genomic  
clone UUGCIM0283A09 F, DNA sequence.

AZ469557  
GSS.  
GI:10627682

house mouse.  
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 23)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0283 row: A column: 09  
Seq primer: CGTTGTAACAGCGGCCAGT  
Class: plasmid ends  
High quality sequence stop: 23.  
Location/Qualifiers

FEATURES  
source

```
1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCLM0283A09"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
```

BASE COUNT 0 a 9 c 2 g 12 t  
ORIGIN

Query Match 50.8%; Score 12.2; DB 17; Length 23;  
Best Local Similarity 82.4%; Pred. No. 3.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGACGACGAAGACACA 17  
||||| ||| |||||  
DB 19 GAGAGAGAGAGACACA 3

RESULT 2  
AZ434508 23 bp DNA linear GSS 03-OCT-2000  
LOCUS  
DEFINITION IM0220K22R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0220K22 R, DNA sequence.

ACCESSION AZ434508.1 GI:10558521  
VERSION  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 23)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00

Plate: 0220 row: K column: 22  
Seq primer: CACACAGGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 23.  
Location/Qualifiers

FEATURES  
source

```
1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCLM0220K22"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
```

BASE COUNT 9 a 8 c 4 g 2 t  
ORIGIN

Query Match 48.3%; Score 11.6; DB 17; Length 23;  
Best Local Similarity 77.8%; Pred. No. 5.6e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 AAGAACGACGACGACGCC 24  
||||| ||||| |||||  
DB 3 AAGGCACGACGACGACCC 20

RESULT 3  
AZ937666/c 22 bp DNA linear GSS 26-APR-2001  
LOCUS  
DEFINITION 2M0195L23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0195L23 R, DNA sequence.

ACCESSION AZ937666  
VERSION  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 22)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00

Plate: 0195 row: L column: 23  
 Seq primer: CACAGGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 22.

## FEATURES

Location/Qualifiers  
 1. .22  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG2M0195L23"  
 /clone\_lib="Mouse 10kb plasmid UUGC2M library"  
 /sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

2 a 11 c 2 g 7 t

## BASE COUNT

Query Match  
 Best Local Similarity 47.5%; Score 11.4; DB 17; Length 22;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGAGCAAGAAAG 13  
 |||||

Db 22 GAGAGCAAGCAAG 10  
 |||||

## RESULT 4

TA372G12P 22 bp DNA linear GSS 13-DEC-2000  
 LOCUS T. brucei sheared genomic DNA clone 372g12, forward sequence,  
 DEFINITION genomic survey sequence.  
 ACCESSION AL496138  
 VERSION AL496138.1 GI:11872177  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma  
 1 (bases 1 to 22)

REFERENCE  
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.

TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhle@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

## FEATURES

source

Location/Qualifiers  
 1. .22  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="372g12"

BASE COUNT 11 a 3 c 8 g 0 t

## ORIGIN

Query Match 47.5%; Score 11.4; DB 17; Length 22;  
 Best Local Similarity 71.4%; Pred. No. 6.6e+05;  
 Matches 15; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 AGAGCAAGAAAGCACAGCAAG 22  
 |||

Db 2 AGCGAAAGGAGCGAAGAAAG 22  
 |||

## RESULT 5

AZ486363/c

LOCUS

DEFINITION 1M0314015F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0314015 F, DNA sequence.

ACCESSION AZ486363

VERSION AZ486363.1 GI:10653065

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 23)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beccorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0314 row: 0 column: 15  
 Seq primer: CGTTGTAACGACGCGCAGT  
 Class: plasmid ends  
 High quality sequence stop: 23.

## FEATURES

source

Location/Qualifiers  
 1. .23  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0314015"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

Barrell, Oxford University Press, 1999).  
 Email: nelsaved@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/T-brucei/.

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 4 c 7 g 7 t  
ORIGIN  
Query Match 47.5%; Score 11.4; DB 17; Length 23;  
Best Local Similarity 92.3%; Pred. NO. 6.7e+05;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 AAGCAGCAGCAGC 23  
|||||||  
Db 21 AAGCAGCAGCAGC 9  
|||||||

RESULT 6  
AZ648037/c  
LOCUS  
DEFINITION  
LM0514C23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0514C23 R, DNA sequence.

ACCESSION  
AZ648037  
VERSION  
AZ648037.1 GI:11780102  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 25)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0514 row: C column: 23  
Seq primer: CACACAGGAACAGCATGACC  
Class: plasmid ends  
High quality sequence stop: 25.  
Location/Qualifiers  
1. .25  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0514C23"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

JOURNAL  
COMMENT

TITLE

JOURNAL

COMMENT

FEATURES  
source

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 0 a 6 c 9 g 10 t  
ORIGIN  
Query Match 47.5%; Score 11.4; DB 17; Length 25;  
Best Local Similarity 92.3%; Pred. NO. 6.8e+05;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 AGAAGCAGCAGCA 20  
|||||||  
Db 16 AGACAGCAGCAGCA 4  
|||||||

RESULT 7  
AZ453505/c  
LOCUS  
DEFINITION  
IM0254M22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0254M22 R, DNA sequence.

ACCESSION  
AZ453505  
VERSION  
AZ453505.1 GI:10611371  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 20)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0254 row: M column: 22  
Seq primer: CACACAGGAACAGCATGACC  
Class: plasmid ends  
High quality sequence stop: 20.  
Location/Qualifiers  
1. .20  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0254M22"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

JOURNAL  
COMMENT

TITLE

JOURNAL

COMMENT

FEATURES  
source

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 4 c 2 g 12 t  
ORIGIN

Query Match 46.7%; Score 11.2; DB 17; Length 20;  
Best Local Similarity 81.2%; Pred. No. 7.7e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAGCACA 17  
||||| |||| |||  
DB 19 AGAGCATGAAGAGCA 4

## RESULT 8

AZ357282/c

LOCUS

DEFINITION 1M0098A16R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0098A16 R, DNA sequence.

ACCESSION AZ357282

VERSION AZ357282.1

KEYWORDS GI:10470982

SOURCE GSS.

ORGANISM

house mouse.

MUS MUSCULUS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 23)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

Plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0098 row: A column: 16

Seq primer: CACACAGAACACGTATGACC

Class: plasmid ends

High quality sequence stop: 23.

Location/Qualifiers

1. .23

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0098A16"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

## FEATURES

source

1. .24

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0286N24"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 4 c 3 g 14 t  
ORIGIN

Query Match 46.7%; Score 11.2; DB 17; Length 23;  
Best Local Similarity 81.2%; Pred. No. 8e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 AAGAAACACAGCAAG 22  
||||| |||| |||  
DB 16 AAGAAACACACAAAG 1

## RESULT 9

AZ999558/c

LOCUS

DEFINITION 2M0286N24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0286N24 R, DNA sequence.

ACCESSION AZ999558

VERSION AZ999558.1

KEYWORDS GI:13870785

SOURCE GSS.

house mouse.

MUS MUSCULUS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 24)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

Plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0286 row: N column: 24

Seq primer: CACACAGAACACGTATGACC

Class: plasmid ends

High quality sequence stop: 24.

Location/Qualifiers

1. .24

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0286N24"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 11.0 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1147321141gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

DATE	TIME	PLACE	COUNT	REMARKS
10/10/66	10:00	4 a	6 c	
10/10/66	10:00	3 q	6 t	

Query Match 45.8%; Score 11; DB 17; Length 19;  
Best Local Similarity 100.0%; Pred. No. 9.1e+05;  
Matches 11: Conservative 0; Mismatches 0; Indels

QY 4 AGCAAGAAAGC 14  
          |||||  
Db 18 AGCAAGAAAGC 8

RESULT 11	20 bp	DNA	linear	GSS 20-FEB-2001
Az834080/c		plasmid		
LOCUS	AZ834080	2M0116A09	Mouse 10kb	library Mmus musculus genomic
DEFINITION				class: mrc22m0116a09.p DNA sequence.

AZ834080  
 AZ834080.1  
 GI:13003988  
 GSS.

house mouse.  
SOURCE  
Mus musculus.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 20)  
Dunn D, Aoyagi A., Barber M., Beacorn T., Duval B., Hamil C.,  
Islam H., Longacre S., Mahmoud M., Meenen E., Pedersen T., Reilly  
M., Rose M., Rose R., Stokes R., Tingey A., von Niederhausern A.  
and Wright D., Weiss R.  
TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb

TITLE	Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts
1	Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts

**JOURNAL  
COMMENT**

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: [gdunn@genetics.utah.edu](mailto:gdunn@genetics.utah.edu)  
 Err: 801 585 7177  
 Insert Length: 10000 Std Err: 10000  
 Plate: 0116 row: A column: 0  
 Seq primer: CACACAGGACACTAT  
 Class: plasmid ends  
 High quality sequence stop: 20

FEATURES  
SOURCE

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0116A03"
/clone_lib="Mouse 10kb plasmid UUCGIM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42mv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

```

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 1 a 4 c 4 g 11 t  
ORIGIN

Query Match 45.8%; Score 11; DB 17; Length 20;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAA 12  
|||||  
Db 11 AGAGCAAGAAA 1

RESULT 12  
AZ778465/c  
LOCUS  
DEFINITION  
2M0013G17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0013G17 R, DNA sequence.  
ACCESSION  
AZ778465  
VERSION  
AZ778465.1 GI:12908136  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 24)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.,  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0013 row: G column: 17  
Seq primer: CACACAGGAACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 24.  
Location/Qualifiers  
1. .24

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0013G17"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The

FEATURES  
source

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 6 c 5 g 10 t  
ORIGIN

Query Match 45.8%; Score 11; DB 17; Length 24;  
Best Local Similarity 73.7%; Pred. No. 9.8e+05;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAACGACAGCA 20  
|||||  
Db 21 AGAGCAAGAGCTCACTGCA 3

RESULT 13  
AZ796553  
LOCUS  
DEFINITION  
2M0052P15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0052P15 F, DNA sequence.  
ACCESSION  
AZ796553  
VERSION  
AZ796553.1 GI:12944728  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 20)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.,  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0052 row: P column: 15  
Seq primer: CGTGTAAACGACGCCACGT  
Class: plasmid ends  
High quality sequence stop: 20.  
Location/Qualifiers  
1. .20

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0052P15"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The

FEATURES  
source

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 11 a 2 c 6 g 1 t  
ORIGIN

Query Match 45.0%; Score 10.8; DB 17; Length 20;  
Best Local Similarity 85.7%; Pred. No. 1.1e+06;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGACAAGAAGCA 15  
||||| ||||| |||  
Db 6 AGAGAAGAATGCA 19

RESULT 14  
AZ808800/c  
LOCUS 20 bp DNA linear GSS 20-FEB-2001  
DEFINITION 2M0072F01R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGC2M0072F01 R, DNA sequence.

ACCESSION AZ808800  
VERSION  
KEYWORDS  
SOURCE GSS.  
ORGANISM house mouse.

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 20)  
REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R., Stokes, R., Tingey, A., von Niederhausern, A.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0072 row: F column: 01  
Seq primer: CACACGAAACACGATGACC  
Class: plasmid ends  
High quality sequence stop: 20.  
Location/Qualifiers  
1. 20  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0072F01"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

FEATURES  
source  
1. 20  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0072F01"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 1 a 8 c 3 g 8 t  
ORIGIN

Query Match 45.0%; Score 10.8; DB 17; Length 20;  
Best Local Similarity 85.7%; Pred. No. 1.1e+06;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCAAGAAAGCAGACAG 18  
||||| ||||| |||  
Db 20 GCAAGAAAGCAGAAAG 7

RESULT 15  
AZ308814/c  
LOCUS 21 bp DNA linear GSS 29-SEP-2000  
DEFINITION 1M0012N07F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0012N07 F, DNA sequence.

ACCESSION AZ308814  
VERSION  
KEYWORDS  
SOURCE GSS.  
ORGANISM house mouse.

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 21)  
REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R., Stokes, R., Tingey, A., von Niederhausern, A.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0012 row: N column: 07  
Seq primer: CGTGTAAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 21.  
Location/Qualifiers  
1. 21  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0012N07"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

FEATURES  
source  
1. 21  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0012N07"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The



adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil473211419b/Ar129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	3 a	8 c	2 g	8 t
ORIGIN				
Query Match	44.2%; Score 10.6; DB 17; Length 21;			
Best Local Similarity	76.5%; Pred. No. 1.4e+06;			
Matches	13; Conservative	0; Mismatches	4; Indels	0; Gaps 0;
QY	2 AGAGCAAGAAACACAG 18			
Db				
	17 AGAGTAAGTACCGGAG 1			

Search completed: July 17, 2003, 22:19:04  
Job time : 1646 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:01:47 ; Search time 48 Seconds  
(without alignments)  
153.338 Million cell updates/sec

Title: US-09-923-684-12  
Perfect score: 24  
Sequence: 1 gagagcaagaagcacagcagcc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 364038

Minimum DB seq length: 0  
Maximum DB seq length: 25

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_patents\_NA:.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13.6	56.7	22	4	US-09-467-997-7
2	13.4	55.8	24	2	US-08-117-952-239
3	13.2	55.0	22	2	US-08-634-797-9
4	13.2	55.0	23	1	US-08-406-635-10
5	13	54.2	21	1	US-08-263-413-21
6	12.8	53.3	21	2	US-08-117-952-295
7	12.6	52.5	23	4	US-09-338-907-143
8	12.6	52.5	23	4	US-09-218-207-143
9	12.4	51.7	19	4	US-09-147-208-70
10	12.4	51.7	25	4	US-09-315-794-8
11	12.4	51.7	25	4	US-09-389-341-8
12	12.2	50.8	18	3	US-09-339-964-46
13	12.2	50.8	20	1	US-08-729-447-15
14	12.2	50.8	20	2	US-08-715-890-1
15	12.2	50.8	20	2	US-08-743-637B-159
16	12.2	50.8	20	3	US-08-526-840B-159
17	12.2	50.8	22	1	US-08-729-447-12
18	12.2	50.8	25	1	US-08-410-804-11
19	12.2	50.8	25	1	US-08-607-269-14
20	12.2	50.8	25	1	US-08-259-514-11
21	12.2	50.8	25	2	US-08-858-311-11
22	12.2	50.8	25	4	US-08-697-610-7
23	12.2	50.8	25	4	US-08-349-357-7
24	12.2	50.8	25	5	PCT-US95-04600-14
25	12	50.0	19	1	US-08-268-799-4
26	12	50.0	19	6	5166195-2
27	12	50.0	20	1	US-08-308-869-2

C 28	12	50.0	20	1	US-08-782-980-2	Sequence 2, Appli
C 29	12	50.0	20	2	US-08-808-474A-19	Sequence 19, Appl
C 30	12	50.0	20	2	US-08-808-474A-20	Sequence 20, Appl
C 31	12	50.0	20	4	US-09-235-614-19	Sequence 19, Appl
C 32	12	50.0	20	4	US-09-235-614-20	Sequence 20, Appl
C 33	12	50.0	20	4	US-09-517-584A-63	Sequence 63, Appl
C 34	12	50.0	20	4	US-09-036-637D-2	Sequence 2, Appli
C 35	12	50.0	20	4	US-09-036-637D-5	Sequence 5, Appli
C 36	12	50.0	21	6	5166195-3	Patent No. 5166195
C 37	12	50.0	22	4	US-09-092-077-20	Sequence 20, Appl
C 38	12	50.0	24	1	US-08-412-229-8	Sequence 8, Appli
C 39	12	50.0	24	1	US-08-825-617-8	Sequence 8, Appli
40	11.8	49.2	20	1	US-08-602-203-1	Sequence 1, Appli
41	11.8	49.2	20	3	US-09-009-913-331	Sequence 331, App
C 42	11.8	49.2	20	4	US-09-517-467B-303	Sequence 303, App
C 43	11.8	49.2	23	2	US-08-687-865A-12	Sequence 12, Appl
44	11.8	49.2	23	4	US-09-043-711-12	Sequence 12, Appl
C 45	11.8	49.2	23	4	US-09-240-918-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1  
US-09-467-997-7/c  
; Sequence 7, Application US/09467997  
; Patent No. 6379925  
; GENERAL INFORMATION:  
; APPLICANT: Kitajewski, Jan  
; APPLICANT: Uytendaele, Hendrik  
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION  
; FILE REFERENCE: 53863-A-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/467,997  
; CURRENT FILING DATE: 1999-12-20  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: mouse  
US-09-467-997-7

Query Match 56.7%; Score 13.6; DB 4; Length 22;  
Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAAGAAAGCACAGCAGCC 24  
DB 21 GCAAGAAAGCACAGCAGC 2

RESULT 2  
US-08-117-952-239  
; Sequence 239, Application US/08117952  
; Patent No. 5851760  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Smith, Michael W.  
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE  
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES  
; NUMBER OF SEQUENCES: 797  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```
;
; MOLECULE TYPE: DNA (genomic)
; US-08-634-797-9
;
; Query Match 55.0%; Score 13.2; DB 2; Length 22;
; Best Local Similarity 83.3%; Pred. No. 1.7e+03;
; Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; QY 4 AGCAAGAAAGCACAGCAA 21
;      ||||| |||||
; Db 18 AGCAAGAAAGCACAGCAA 1
;
; RESULT 4
; US-08-406-635-10/c
; Sequence 10, Application US/08406635
; Patent No. 559674
; GENERAL INFORMATION:
; APPLICANT: PENA, SERGIO D.J.
; APPLICANT: SIMPSON, ANDREW J.G.
; TITLE OF INVENTION: METHOD FOR RECOGNITION OF THE NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE OF A PURIFIED DNA SEGMENT
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,635
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,738
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 45119-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-406-635-10
;
; Query Match 55.0%; Score 13.2; DB 1; Length 23;
; Best Local Similarity 83.3%; Pred. No. 1.7e+03;
; Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; QY 6 CAAGAAAGCACAGCAAGC 23
;      ||||| |||||
; Db 19 CAAGCAAGTACAGCAATC 2
;
; RESULT 5
; US-08-263-413-21
; Sequence 21, Application US/08263413
; Patent No. 5747246
; GENERAL INFORMATION:
; APPLICANT: PANNETIER, CHRISTOPHE
; APPLICANT: COCHET, MADELEINE
```

APPLICANT: DANCHE, SYLVIE  
APPLICANT: KOURILSKY, PHILIPPE  
TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF  
TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF  
TITLE OF INVENTION: METHOD OF ENZYMIC AMPLIFICATION OF DNA  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/263,413  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,980  
FILING DATE: 14-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, J.P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 660-058-55X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-263-413-21

Query Match 54.28; Score 13; DB 1; Length 21;  
Best Local Similarity 76.2; Pred. No. 2e+03;  
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCAGCAGG 24  
Db 1 AACAAAGAAAGCAGCAGG 21

## RESULT '6

US-08-117-952-295/c  
Sequence 295, Application US/08117952  
Patent No. 5851760  
GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
APPLICANT: Smith, Michael W.  
TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE  
TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES  
NUMBER OF SEQUENCES: 797  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/117,952  
FILING DATE: 07-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/078,471  
FILING DATE: 15-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 295:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Oligonucleotide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-117-952-295

Query Match 53.3%; Score 12.8; DB 2; Length 21;  
Best Local Similarity 87.5%; Pred. No. 2.5e+03;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAGCAAGAAAGCAGCAG 18  
Db 16 GAACATGAAGCAGCAG 1

## RESULT 7

US-09-338-907-143  
Sequence 143, Application US/09338907  
Patent No. 6265546  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilyu, Chumakov  
APPLICANT: Bouqueleret, Lydie  
TITLE OF INVENTION: PROSTATE CANCER GENE  
FILE REFERENCE: GENSET.18CPICP  
CURRENT APPLICATION NUMBER: US/09/338,907  
CURRENT FILING DATE: 1999-06-23  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 09/218,207  
EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 143  
LENGTH: 23  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..23  
OTHER INFORMATION: amplification oligonucleotide pGIASe24  
US-09-338-907-143

Query Match 52.5%; Score 12.6; DB 4; Length 23;  
Best Local Similarity 78.9%; Pred. No. 3.1e+03;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCAGCAGG 22  
Db 2 ATCATCAAGCAGCAGCATG 20

```

; LENGTH: 19 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-147-208-70

Query Match          51.7%; Score 12.4; DB 4; Length 19;
Best Local Similarity 92.9%; Pred. No. 3.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      9 GAAAGCACAGCAAG 22
Db      18 GAAAGGACAGCAAG 5
        ||||| |||||

RESULT 10
US-09-315-794-8
; Sequence 8, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
; APPLICANT: Robert's, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-053
; CURRENT APPLICATION NUMBER: US/09/315,794
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
; US-09-315-794-8

Query Match          51.7%; Score 12.4; DB 4; Length 25;
Best Local Similarity 92.9%; Pred. No. 3.8e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 ACCAAGAAAGCACCA 17
Db      7 AGAAGAAAGCACCA 20
        || ||||| |||||

RESULT 11
US-09-389-341-8
; Sequence 8, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
; APPLICANT: Robert's, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-057
; CURRENT APPLICATION NUMBER: US/09/389,341
; CURRENT FILING DATE: 1999-09-02
; EARLIER APPLICATION NUMBER: 09/315,794
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
; US-09-389-341-8

Query Match          51.7%; Score 12.4; DB 4; Length 25;
Best Local Similarity 92.9%; Pred. No. 3.8e+03;
```

```

; LENGTH: 23
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..23
; OTHER INFORMATION: amplification oligonucleotide PGLASe24
; US-09-218-207-143

Query Match          52.5%; Score 12.6; DB 4; Length 23;
Best Local Similarity 78.9%; Pred. No. 3.1e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 AGCAAGAAAGCACAGCAAG 22
Db      2 ATCATCAAGACAGCATG 20
        || || ||||| |||||

RESULT 9
US-09-147-208-70/c
; Sequence 70, Application US/09147208
; Patent No. 6333303
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Antiviral Ricin-Like Proteins
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/147,208
; APPLICATION NUMBER: US/09/147,208
; FILING DATE: 02-MAR-1999
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Rudolph, John R.
; REGISTRATION NUMBER: 38,003
; REFERENCE/DOCKET NUMBER: 7841-76
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
```

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCACCA 17  
Db 7 AGAAGAAAGCACCA 20

RESULT 12  
US-09-339-964-46  
; Sequence 46, Application US/09339964  
; Patent No. 6025198  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Lex M. Cowsett  
; TITLE OF INVENTION: ANTISENSE MODULATION OF SHIP-2 EXPRESSION  
; FILE REFERENCE: RTS-0065  
; CURRENT APPLICATION NUMBER: US/09/339,964  
; CURRENT FILING DATE: 1999-06-25  
; NUMBER OF SEQ ID NOS: 47  
; SEQ ID NO 46  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Antisense Oligonucleotide  
US-09-339-964-46

Query Match 50.8%; Score 12.2; DB 3; Length 18;  
Best Local Similarity 82.4%; Pred. No. 4.4e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGCAAGAAAGCACAG 18  
Db 1 AGACGAGGCACACAG 17

RESULT 13  
US-08-729-447-15  
; Sequence 15, Application US/08729447  
; Patent No. 5789174  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: DETECTION OF PERIODONTAL PATHOGENS INCLUDING  
; TITLE OF INVENTION: BACTEROIDES FORSYTHUS, PORPHYROMONAS GINGIVALIS,  
; TITLE OF INVENTION: PREVOTELLA INTERMEDIA AND PREVOTELLA NIGRESCENS  
; NUMBER OF SEQUENCES: 20  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA: US/08/729,447  
; FILING DATE:  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "PCR primer pair No. 5789174 3 for  
; DESCRIPTION: P. intermedia"  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-08-729-447-15

Query Match 50.8%; Score 12.2; DB 1; Length 20;  
Best Local Similarity 82.4%; Pred. No. 4.5e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCACCA 20

Db 2 AGCAAGAAAGCAAGGA 18

RESULT 14  
US-08-715-890-1  
; Sequence 1, Application US/08715890  
; Patent No. 5919622  
; GENERAL INFORMATION:  
; APPLICANT: MACHO, Heinz  
; APPLICANT: BEINHAUS, Gerhard  
; TITLE OF INVENTION: SYSTEM FOR THE TEMPERATURE  
; TITLE OF INVENTION: ADJUSTMENT TREATMENT OF LIQUID SAMPLES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIKAYDO, MARCELSTEIN, MURRAY & ORAM LLP  
; STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
; STREET: Street Lobby  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/715,890  
; FILING DATE: 19-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERMAN, Richard J.  
; REGISTRATION NUMBER: 39,107  
; REFERENCE/DOCKET NUMBER: PL614-6054  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202/638-5000  
; TELEFAX: 202/638-4810  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Oligodesoxyribonucleotide"  
; HYPOTHETICAL: NO  
US-08-715-890-1

Query Match 50.8%; Score 12.2; DB 2; Length 20;  
Best Local Similarity 82.4%; Pred. No. 4.5e+03;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 AGAAGACACAGCAAGCC 24  
Db 1 AGACAGTACAGCCAGCC 17

RESULT 15  
US-08-743-637B-159/c  
; Sequence 159, Application US/08743637B  
; Patent No. 5994066  
; GENERAL INFORMATION:  
; APPLICANT: BERGERON, Michel G.  
; APPLICANT: PICARD, Francois J.  
; APPLICANT: OUELLETTE, Marc  
; APPLICANT: ROY, Paul H.  
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:

us-09-923-684-12.sz1m25.rni

Mon Jul 21 15:46:53 2003

```

; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 159:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; US-08-743-637B-159

```

```

Query Match      50.8%; Score 12.2; DB 2; Length 20;
Best Local Similarity 82.4%; Pred. No. 4.5e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 AGAGCAAGAAAGCACAG 18
      ||| ||||| |||
Db      19 AGAACAAAGAAAGGAAAG 3

```

Search completed: July 17, 2003, 21:51:26  
Job time : 51 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 16:48:04 ; Search time 7266 Seconds  
(without alignments)  
11451.269 Million cell updates/sec

Title: US-09-923-684-2  
Perfect score: 2859  
Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2059

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hcg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pi.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_un.\*
- 28: em\_vi.\*
- 29: em\_vt.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_rtd.\*
- 37: em\_htg\_vit.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	2859	100.0	2859	9	HSU80457	U80457 Human trans
2	1661.6	58.1	3921	9	HSU80456	U80456 Human trans
3	1593.6	55.7	136219	9	AP000697	AP000697 Homo sapi
4	1593.6	55.7	340000	9	AP001726	AP001726 Homo sapi
5	1590.4	55.6	3365	9	AB003185	AB003185 Homo sapi
6	1584.8	55.4	120000	2	HSAC000001	AC000001 Homo sapi
7	1389	48.6	1465	9	HS001858	AJ001858 Homo sapi
8	1323.4	46.3	3614	6	E14065	E14065 cDNA encodi
9	1323.4	46.3	3614	10	D64135	D64135 Mus musculu
10	1321.8	46.2	2914	10	MUSMSIM6J	D63383 Mouse mRNA
11	1320.2	46.2	3071	10	MMU42554	D2554 Mus musculu
12	1257.2	44.0	3963	10	MMU40576	U40576 Mus musculu
13	1142.8	40.0	250000	9	AB063285	AB063285 Homo sapi
14	939	32.8	2309	5	AF283298	AF283298 Xenopus l
15	704.6	24.6	4007	9	HSU70212	U70212 Human SIM1
16	671.8	23.5	2547	10	D79209	D79209 Mus musculu
17	665.8	23.3	2664	5	AY028626	AY028626 Danio rer
18	662.8	23.2	4380	5	AF363019	AF363019 Danio rer
19	655.8	22.9	7369	10	MMU40575	U40575 Mus musculu
20	397.6	13.9	2838	3	AV129457	AV129457 Drosophil
21	370.6	13.0	2747	3	DROSIMNP	MI9020 D.melanogas
22	296	10.4	1532	10	MMSM2G09	AF023872 Mus muscu
23	234	8.2	42338	9	HUMQ16F8	D85922 Human DNA f
24	233.8	8.2	147054	2	AL356582	AL356582 Homo sapi
25	232.6	8.1	160810	9	AC009953	AC009953 Homo sapi
26	232.2	8.1	167548	9	AL391809	AL391809 Human DNA
27	231.6	8.1	162771	9	AC027121	AC027121 Homo sapi
28	231.6	8.1	163157	2	AC012567	AC012567 Homo sapi
29	231.6	8.1	166338	9	AC026165	AC026165 Homo sapi
30	231.6	8.1	208879	9	AC024171	AC024171 Homo sapi
31	231.4	8.1	238	6	E14059	E14059 A part of h
32	231.4	8.1	238	9	HUMSIMA	D4444 Homo sapien
33	231.2	8.1	178367	9	AC006115	AC006115 Homo sapi
34	231	8.1	43128	9	AC011558	AC011558 Homo sapi
35	231	8.1	156197	9	HS97K10	281365 Human DNA s
36	230.6	8.1	50000	6	AX393458	AX393458 Sequence
37	230.6	8.1	83549	9	AC004752	AC004752 Homo sapi
38	230.6	8.1	107480	9	AL596094	AL596094 Human DNA
39	230.6	8.1	184268	2	AC011401	AC011401 Homo sapi
40	230.6	8.1	208145	2	AC053504	AC053504 Homo sapi
41	230.4	8.1	3605	5	AF304864	AF304864 Oncorhnc
42	230.4	8.1	137074	9	AC072052	AC072052 Homo sapi
43	230.4	8.1	195342	9	AC009120	AC009120 Homo sapi
44	230	8.0	115231	9	HS785G19	AL035458 Human DNA
45	230	8.0	182753	2	AL137253	AL137253 Homo sapi

ALIGNMENTS

RESULT 1  
LOCUS HSU80457  
DEFINITION Human transcription factor SIM2 short form mRNA linear PRI 08-JUL-1997  
ACCESSION U80457  
VERSION U80457.1 GI:2062418  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2859)  
AUTHORS Chrast,R., Scott,H.S., Chen,H., Kudoh,J., Rossier,C., Minoshima,S.,  
Wang,Y., Shimizu,N. and Antonarakis,S.E.  
TITLE Cloning of two human homologs of the Drosophila single-minded gene

SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome  
chromosomal region  
Genome Res. 7 (6), 615-624 (1997)  
97343329  
PUBMED  
9199934  
REFERENCE  
AUTHORS  
2 (bases 1 to 2859)  
Chrast,R., Kudoh,J., Rossier,C., Chen,H., Minoshima,S., Shimizu,N.  
and Antonarakis,S.E.  
Direct Submission  
Submitted (29-NOV-1996) Medical Genetics, University of Geneva  
Medical School, 1, Rue Michel-Servet, Geneva 1211, Switzerland  
Location/Qualifiers  
1..2859  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.2"  
93..1805  
/codon\_start=1  
/product="transcription factor SIM2 short form"  
/protein\_id="AAB62397.1"  
/db\_xref="GI:2062419"  
/translation="MKESKNAKTRREKENGFEYELAKLLPLPSAITSOLDKASIIIR  
LITSLKMRVPEGLGDWQPSRAGLDGKVELGSHLIQTLDTGDFVVFVSDGKIM  
YISFASVHLGSLQVELTNGSIYVILHPSDHEMTAVITAHOPHLHLLOREIEREF  
FLRMKCVLAKRNLGTCGKVIHCSGYLKIROYMLDMSLYDSCQIVLGHVAGDSLP  
PSAITEIKLYSNMFRLASLDLKLFLDSRVTEVIGTEPQDLIEKTLXHHVGDVPH  
LRYAHLLLVQVQVTKYIRLLSRGWWVQSVATVHHNSRRSRPHCIYSVNVYLE  
FYKELQSLQGVSTAKQSDSWRTALSTQETRLVKNPKMTKTKURTNPYPQQYS  
SPMDKLECGGLGNWRASPPASAAAPPELOPHSBSLDLLTPSYSLFSPYHGFPLD  
SHFSSKPKMLPAKQPGQSPCEVAREFLSTMPASGBCOWHYANPLVPSYSSSPAKNP  
PEPPANTAHSLVPSYEGSGLLVKGVLRTAGSRSSHHGGWQMETEPFRFGTCPL  
SASK"  
BASE COUNT 701 a 803 c 778 g 577 t  
ORIGIN  
Query Match 100.0%; Score 2859; DB 9; Length 2859;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTGGGGCTCCGGGGCTCGAGCAGC 60  
DB 1 ACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTGGGGCTCCGGGGCTCGAGCAGC 60  
QY 61 GCGGGTCTAATATGCGGAGCGCGGCGGCGATGAAGGAGAGTCCAAAGATGGCGCA 120  
DB 61 GCGGGTCTAATATGCGGAGCGCGGCGGCGATGAAGGAGAGTCCAAAGATGGCGCA 120  
QY 121 AGACGAGGAGGAGGAGGAGGAGTGGCGAGTTTACGAGCTTGCCAGCTGCTCCGCTGC 180  
DB 121 AGACGAGGAGGAGGAGGAGGAGTGGCGAGTTTACGAGCTTGCCAGCTGCTCCGCTGC 180  
QY 181 CGTGGCCATCACTTCCAGCTGCAGAACGCTCCATCATCCGCTTCACACAGCTACC 240  
DB 181 CGTGGCCATCACTTCCAGCTGCAGAACGCTCCATCATCCGCTTCACACAGCTACC 240  
QY 241 TGAAGATGCGCGCGCTCTCCCGGAGGTTTAGGAGACGCGTGGGACAGCGCGCGG 300  
DB 241 TGAAGATGCGCGCGCTCTCCCGGAGGTTTAGGAGACGCGTGGGACAGCGCGCGG 300  
QY 301 CCGGCGCGCTGGAGCGGCTGCCAAGGAGCTGGGATGCGACTTGCTCGAGCTTTGGATG 360  
DB 301 CCGGCGCGCTGGAGCGGCTGCCAAGGAGCTGGGATGCGACTTGCTCGAGCTTTGGATG 360  
QY 361 GATTGTGTTTGTGTAGCATCTGATGCAAAATCATGTATATATCCGAGACCGCTTCTG 420  
DB 361 GATTGTGTTTGTGTAGCATCTGATGCAAAATCATGTATATATCCGAGACCGCTTCTG 420  
QY 421 TCCATTAGGCTTATCCAGGTGAGCTCAGCGGCAACAGTATTATGATACATCATC 480  
DB 421 TCCATTAGGCTTATCCAGGTGAGCTCAGCGGCAACAGTATTATGATACATCATC 480  
QY 481 CTTCGTGACACGATGAGATGACCGCTGTCTTCAGCGGCCACACCGCGCTGCACACC 540

481 CTTCGTGACACGATGAGATGACCGCTGTCTTCAGCGGCCACACCGCGCTGCACACC 540  
QY 541 TGCTCCAAAGAGTATGAGATGAGAGGCTGTTCTTTCGAATGAATGTTCTTTGGCGA 600  
DB 541 TGCTCCAAAGAGTATGAGATGAGAGGCTGTTCTTTCGAATGAATGTTCTTTGGCGA 600  
QY 601 AAAGGAACGCGGCTGACCTGCAGCGGATCAAAAGGTCACTGAGTGGCTACTTGA 660  
DB 601 AAAGGAACGCGGCTGACCTGCAGCGGATCAAAAGGTCACTGAGTGGCTACTTGA 660  
QY 661 AGATCAGGCAAGTATGCTGGACATGTCCTGTAGACTCTCTGCTACAGATTGTGGGC 720  
DB 661 AGATCAGGCAAGTATGCTGGACATGTCCTGTAGACTCTCTGCTACAGATTGTGGGC 720  
QY 721 TGGTGGCGCTGGCGCAGTGGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780  
DB 721 TGGTGGCGCTGGCGCAGTGGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780  
QY 781 ACATGTTTCATGTTTCAGGGCCAGCTTGACCTGAAGTATATCTGGATTCCAGGGTGA 840  
DB 781 ACATGTTTCATGTTTCAGGGCCAGCTTGACCTGAAGTATATCTGGATTCCAGGGTGA 840  
QY 841 CCAGGTGACGGGTTACGAGCCGACGACTGATCGAGAAGACCTTATACCATCATCGTC 900  
DB 841 CCAGGTGACGGGTTACGAGCCGACGACTGATCGAGAAGACCTTATACCATCATCGTC 900  
QY 901 ACGGCTGCGACGTTTCCACCTCCGCTAGCAGCACACACCTCTGTTGGTGAAGGGCCAG 960  
DB 901 ACGGCTGCGACGTTTCCACCTCCGCTAGCAGCACACACCTCTGTTGGTGAAGGGCCAG 960  
QY 961 TCACCCAAAGTACTACCGGCTGCTGCTCAAGCGGGCGGCTGGGTGGTGGTGCAGAGCT 1020  
DB 961 TCACCCAAAGTACTACCGGCTGCTGCTCAAGCGGGCGGCTGGGTGGTGGTGCAGAGCT 1020  
QY 1021 ACGCCACCGTGTGCAACACAGCGCTGCTGCCGGGCCACTGCATCGTGGTGTCAATT 1080  
DB 1021 ACGCCACCGTGTGCAACACAGCGCTGCTGCCGGGCCACTGCATCGTGGTGTCAATT 1080  
QY 1081 ATGACTACAGGAGATTGAATACAGGAAGTTCAGCTGCTCCCTGGAGCAGCTGTCACATG 1140  
DB 1081 ATGACTACAGGAGATTGAATACAGGAAGTTCAGCTGCTCCCTGGAGCAGCTGTCACATG 1140  
QY 1141 CCAAGTCCAGGACTCTCGGAGGACCGCTTGTCTACCTCACAAGAACTAGGAATTAG 1200  
DB 1141 CCAAGTCCAGGACTCTCGGAGGACCGCTTGTCTACCTCACAAGAACTAGGAATTAG 1200  
QY 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTACCCCCACAGC 1260  
DB 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTACCCCCACAGC 1260  
QY 1261 AATACAGCTCATTCCTCAATGGCAAACTTGAATGGCCAGCTCGGAACTGGAGAGCCA 1320  
DB 1261 AATACAGCTCATTCCTCAATGGCAAACTTGAATGGCCAGCTCGGAACTGGAGAGCCA 1320  
QY 1321 GTCCCGCTGCAAGCGCTGCTGCTCCTCCAGAACTGCAGCCCGCTCAGAAAGCAGTGACC 1380  
DB 1321 GTCCCGCTGCAAGCGCTGCTGCTCCTCCAGAACTGCAGCCCGCTCAGAAAGCAGTGACC 1380  
QY 1381 TTCTGTACACGCCATCTTACAGCTGCGCTTCTCTACCATTTAGGACACTTCCCTCTGG 1440  
DB 1381 TTCTGTACACGCCATCTTACAGCTGCGCTTCTCTACCATTTAGGACACTTCCCTCTGG 1440  
QY 1441 ACTCTCACTTCTCAGCAGCAAAAGCAATGTTGCCGGCCAAAGTTCGGCAGGCCCAAG 1500  
DB 1441 ACTCTCACTTCTCAGCAGCAAAAGCAATGTTGCCGGCCAAAGTTCGGCAGGCCCAAG 1500  
QY 1501 GATCCCTTGTGAGGTGGCAGCTTTTCTCAGCACAATGCCAGCCAGCGGTGAATGCC 1560  
DB 1501 GATCCCTTGTGAGGTGGCAGCTTTTCTCAGCACAATGCCAGCCAGCGGTGAATGCC 1560  
QY 1561 AGTGGCATTTATGCCAAACCCCTTAGTGCCTTAGCAGCTGCTCTCCAGTAAAAATCTCCAG 1620  
DB 1561 AGTGGCATTTATGCCAAACCCCTTAGTGCCTTAGCAGCTGCTCTCCAGTAAAAATCTCCAG 1620

Db	1561	AGTGGCATTTATGCCAACCCCTAGTGCCTAGCAGCTGCTCTCCAGCTAAATAATCTCCAG	1620
QY	1621	AGCCACCGGCAACACTGCTAGCACAGCCTGGTGCCAAAGCTACGAAGTGGGTGAGTGC	1680
Db	1621	AGCCACCGGCGAACAATGCTAGGCACAGCCTGGTGCCAAAGCTACGAAGTGGGTGAGTGC	1680
QY	1681	TGCTCGTGGGGAAGTGGGAGGACTGCGCACGCGCGGAGCGAAGCAGCCATGGCGGTG	1740
Db	1681	TGCTCGTGGGGAAGTGGGAGGACTGCGCACGCGCGGAGCGAAGCAGCCATGGCGGTG	1740
QY	1741	GGTGGCAGATGGAGACAGAACCCCTCACGGCTTTGGGCAAACTTGCCTCTCTTCTGCTTCTA	1800
Db	1741	GGTGGCAGATGGAGACAGAACCCCTCACGGCTTTGGGCAAACTTGCCTCTCTTCTGCTTCTA	1800
QY	1801	AGTAGGCTTGTGCTTCTTCTTCTCAATGCAGGTGCTCTCGAGAGTGAGAAATGG	1860
Db	1801	AGTAGGCTTGTGCTTCTTCTTCTCAATGCAGGTGCTCTCGAGAGTGAGAAATGG	1860
QY	1861	CAGTCTGCTGCTCGGCGACACTAGTGACAGTATAAAGGGGCAAGGAAACCCGAGTATC	1920
Db	1861	CAGTCTGCTGCTCGGCGACACTAGTGACAGTATAAAGGGGCAAGGAAACCCGAGTATC	1920
QY	1921	TGGCTTTCACGTAATCTCTGGCCACATTCACCAACCAAGGGGACAGTATTTCAAAA	1980
Db	1921	TGGCTTTCACGTAATCTCTGGCCACATTCACCAACCAAGGGGACAGTATTTCAAAA	1980
QY	1981	CCAGCTCCCATGCTGTGAAACACCCAGCTGCTTCTTTTGCAGATTCTTTTCCACT	2040
Db	1981	CCAGCTCCCATGCTGTGAAACACCCAGCTGCTTCTTTTGCAGATTCTTTTCCACT	2040
QY	2041	CCAAACAGAGTGAATTTTGAGACAAACGGCCTATTGGCTATTTCCTCCAGTATTT	2100
Db	2041	CCAAACAGAGTGAATTTTGAGACAAACGGCCTATTGGCTATTTCCTCCAGTATTT	2100
QY	2101	GGAAGTGGGGAACATATGTTGGAATTTTGGCTTGGGACAGAAATGCCACTACCA	2160
Db	2101	GGAAGTGGGGAACATATGTTGGAATTTTGGGCTTGGGACAGAAATGCCACTACCA	2160
QY	2161	ACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTTCCACTGTAATATGGGTG	2220
Db	2161	ACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTTCCACTGTAATATGGGTG	2220
QY	2221	AAGCTGTATGCTGCTACTCCCAAGGACACACAGTAGGACCTGCGCTGTACATG	2280
Db	2221	AAGCTGTATGCTGCTACTCCCAAGGACACACAGTAGGACCTGCGCTGTACATG	2280
QY	2281	CTAGTTCAACAGAAAGGAATGGCTTTTACCTTCTCTGTTGGCAGGCAAGCAGATGCC	2340
Db	2281	CTAGTTCAACAGAAAGGAATGGCTTTTACCTTCTCTGTTGGCAGGCAAGCAGATGCC	2340
QY	2341	TCTGGGAGATACCGCTAGCTCCCGAGCGAGCTGACTCTCTGTTTGTCTGCTGAC	2400
Db	2341	TCTGGGAGATACCGCTAGCTCCCGAGCGAGCTGACTCTCTGTTTGTCTGCTGAC	2400
QY	2401	AACCCAGGCAAGAGTGGAGGTGGGAACAGAGTTTACCTGCAGACATGATTCCCA	2460
Db	2401	AACCCAGGCAAGAGTGGAGGTGGGAACAGAGTTTACCTGCAGACATGATTCCCA	2460
QY	2461	TTGCACCGACCTTAATCGCGCTCAGAGTGTAGACCGAGATGGTGACAGTGCCTGAGT	2520
Db	2461	TTGCACCGACCTTAATCGCGCTCAGAGTGTAGACCGAGATGGTGACAGTGCCTGAGT	2520
QY	2521	GCCATTAAATGTGGGTGAAGTGCATCAGGATATTATGTCGCCAGCGCGGCTCAGTG	2580
Db	2521	GCCATTAAATGTGGGTGAAGTGCATCAGGATATTATGTCGCCAGCGCGGCTCAGTG	2580
QY	2581	CTCACACTGTAATCCAGCAGCTTTGGGAGGCCAAGTGGGGGAGTCACTGAGTGCAG	2640
Db	2581	CTCACACTGTAATCCAGCAGCTTTGGGAGGCCAAGTGGGGGAGTCACTGAGTGCAG	2640
QY	2641	AGTTTGGCAGACGCTGCCAAGCTGAACCCCATCTCCACTCAAAAATACAAAATTA	2700
Db	2641	AGTTTGGCAGACGCTGCCAAGCTGAACCCCATCTCCACTCAAAAATACAAAATTA	2700
QY	2701	GTGGGCATGCTGGTGAGCACCTCTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGAT	2760
Db	2701	GTGGGCATGCTGGTGAGCACCTCTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGAT	2760
QY	2761	CACCTTGAACCCGGAGGTGGAGGTTCAGTTCAGTAAAGTACATCACTCACTCCAGCC	2820
Db	2761	CACCTTGAACCCGGAGGTGGAGGTTCAGTTCAGTAAAGTACATCACTCACTCCAGCC	2820
QY	2821	TGGCTAACAGAGTGCAGCTGTCTCAAAAAA	2859
Db	2821	TGGCTAACAGAGTGCAGCTGTCTCAAAAAA	2859
RESULT 2			
HSU80456			
LOCUS	HSU80456	3921 bp	mRNA linear PRI 08-JUL-1997
DEFINITION	Human transcription factor SIM2 long form mRNA, complete cds.		
ACCESSION	U80456		
VERSION	U80456.1	GI:2062416	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 3921)		
JOURNAL	Chen, R., Shimizu, N., and Antonarakis, S.E.		
MEDLINE	Wang, Y., Shimizu, N., and Antonarakis, S.E.		
PUBMED	Cloning of two human homologs of the Drosophila single-minded gene		
AUTHORS	SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome		
TITLE	chromosomal region		
JOURNAL	Genome Res. 7 (6), 615-624 (1997)		
MEDLINE	97343329		
PUBMED	9199934		
AUTHORS	2 (bases 1 to 3921)		
TITLE	Chen, R., Kudoh, J., Rossier, C., Chen, H., Minoshima, S., Shimizu, N.		
JOURNAL	Direct Submission		
FEATURES	Submitted (29-NOV-1996) Medical Genetics, University of Geneva		
source	Medical School, 1, Rue Michel-Servet, Geneva 1211, Switzerland		
CDS	Location/Qualifiers		
	1..3921		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="21"		
	/map="21q22.2"		
	93..2096		
	/codon_start=1		
	/product="transcription factor SIM2 long form"		
	/protein_id="AAB62396.1"		
	/db_xref="GI:2062417"		
	/translation="MKRSKNAKTRREKNGEFYELAKLLPLPSATTSQDKASIIIR		
	LITSTYKRWAVPEGLDAGWOPSRAGLDGVAKEGSHLLQTLDTGDFVVFVSDGKIM		
	YISEFASVHLGSLQVELTNSIYIEHPSDHDEMTAVLTHQPLHLHLLOEYERSF		
	FLRMKCVLAKRNAGLTCGKYVHCQSKIKIROYLMDSLVDSCYQIVGLVAGQSLP		
	PSAITEIKLYSNMFRAKLDLKLIFLDSRVTEVTGPEODLIEKTLHYHGGDVEH		
	LRYAHLLLVKQVTKYRLLSRGHWVQSVATVYVHNSRSRPHCHIVSVNVLTE		
	LEYKELQSLQVSTAKSDSWRTALSTQETRNKLVKPKNTKMTKURNTPPYQOYS		
	SFQDKLSCGQLGNRPSAASAAPELQPHSESSLLYTPSYSLPFSYHYHGFPLD		
	SHVPSKKMPLPAKFGQPGQCEVAREFTSLTPASGECOWHYNPLVPSSSPAKNP		
	PEPPANTARSLVPSYEAAAVRRFGEDAPPFPSCGHYREBPALCPAKAARQAAR		
	DGRLAIAAAPCAECAPTPEAPGAPQLPFVLLNHYRLVLRRLGRLGPAAPASGLAC		
	APGGPEAATGALRLRHPSAATSPGAPLPHYLCASVIITNGR"		
BASE COUNT	922 a 1137 c 998 g 864 t		
ORIGIN			
Query Match	58.1%	Score 1661.6;	DB 9; Length 3921;
Best Local Similarity	99.8%;	Pred. No. 0;	
Matches 1664;	Conservative	0; Mismatches	4; Indels 0; Gaps 0;
QY	1	ACTCATTAGGGCTCGAGCGCGCGCGGCGAGGTGGGCTCGCGGGGCTGGAGCAGC	60
Db	1	ACTCATTAGGGCTCGAGCGCGCGCGGCGAGGTGGGCTCGCGGGGCTGGAGCAGC	60

exon

exon

exon	/note="GENSCAN, score = 3.50, comment = Internal_exon" /evidence-not_experimental 473..576	repeat_region	/rpt_family="MIR" complement(10097..10832) /evidence-not_experimental /rpt_family="HAL1" complement(10251..10334)
repeat_region	/note="GENSCAN, score = 10.54, comment = Terminal_exon" complement(990..1314) /evidence-not_experimental /rpt_family="MER7A" 1228..1233	exon	/note="GRAIL, score = 51.000%, comment = good" complement(11947..12253) /evidence-not_experimental /rpt_family="AluSx" complement(12605..13033)
exon	/note="GENSCAN, score = -0.45, comment = Poly-A_signal" /evidence-not_experimental 1341..1361	repeat_region	/rpt_family="MLT1C" complement(13464..13584) /note="GENSCAN, score = 2.50, comment = Internal_exon" /evidence-not_experimental complement(14406..14623)
repeat_region	/evidence-not_experimental /rpt_family="(CA)n" 1439	exon	/rpt_family="MER20" 14721..14816 /evidence-not_experimental /rpt_family="L2" 14874..15181
exon	/note="GENSCAN, score = 1.05, comment = Poly-A_signal" /evidence-not_experimental 1495..1519	repeat_region	/evidence-not_experimental /rpt_family="AluJo" complement(15887..16217) /evidence-not_experimental /rpt_family="MLT1J2" complement(16153..16192)
repeat_region	/note="GRAIL, score = 78.000%, comment = excellent" /evidence-not_experimental 2023..2072	exon	/note="GENSCAN, score = -0.52, comment = Internal_exon" complement(16355..16468) /note="GENSCAN, score = 3.32, comment = Internal_exon" /evidence-not_experimental complement(16676..16777)
exon	/rpt_family="AT_rich" 3132..3441	repeat_region	/evidence-not_experimental /rpt_family="FLAM_A" complement(16814..17127) /evidence-not_experimental /rpt_family="AluJb" 17287..17313
repeat_region	/evidence-not_experimental /rpt_family="AluJo" 3527..3824	exon	/evidence-not_experimental /rpt_family="AT_rich" complement(17346..17444) /note="GRAIL, score = 51.000%, comment = good" complement(17734..18099)
exon	/evidence-not_experimental /rpt_family="AluJo" complement(4324..4348)	repeat_region	/evidence-not_experimental /rpt_family="L1PB1" 18100..18167 /evidence-not_experimental /rpt_family="(TATATG)n" complement(18170..18443)
repeat_region	/note="GENSCAN, score = -3.5, comment = Terminal_exon" complement(5246..5542) /evidence-not_experimental /rpt_family="AluJb" complement(6054..6168)	exon	/evidence-not_experimental /rpt_family="L1PB1" 18487..18802 /evidence-not_experimental /rpt_family="MLT2FB" 18803..19083
exon	/note="GRAIL, score = 93.000%, comment = excellent" complement(6437..6585) /note="GENSCAN, score = 7.23, comment = Internal_exon" complement(6811..6967)	repeat_region	/evidence-not_experimental /rpt_family="AluSx" 19084..19106 /evidence-not_experimental /rpt_family="(CAAA)n" 19107..19153
repeat_region	/note="GRAIL, score = 63.000%, comment = good" 6829..6960 /evidence-not_experimental /rpt_family="MIR" complement(7095..7134)	exon	/evidence-not_experimental /rpt_family="MLT1A1" complement(8545..8853) /evidence-not_experimental /rpt_family="AluY" 8895..9109
exon	/note="GRAIL, score = 69.000%, comment = good" complement(7227..7352) /evidence-not_experimental /rpt_family="MER90" complement(7468..7514)	repeat_region	/evidence-not_experimental /rpt_family="(TC)n" 19183..19196 /evidence-not_experimental /rpt_family="MLT2FB" 19183..19196
repeat_region	/note="GRAIL, score = 83.000%, comment = excellent" complement(7529..7598) /evidence-not_experimental /rpt_family="MLT1A1" 7599..7867	repeat_region	
exon	/evidence-not_experimental /rpt_family="AluSx" complement(7868..8208) /evidence-not_experimental /rpt_family="MLT1A1" complement(8545..8853)	repeat_region	
repeat_region	/evidence-not_experimental /rpt_family="AluY" 8895..9109 /evidence-not_experimental /rpt_family="MER20" 9834..10061	repeat_region	

[illegible]

## AUTHORS

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, K., Nagamine, K., Mitsuyama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesse, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Mizetic, D., Francis, F., Leirach, H., Reinhardt, R. and Yaspo, M.L.

The DNA sequence of human chromosome 21

Nature 405 (6784), 311-319 (2000)

20289799

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

2 (bases 1 to 340000)

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesse, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Mizetic, D., Francis, F., Leirach, H., Reinhardt, R. and Yaspo, M.L.

## Direct Submission

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Human Genome Research Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717344.

The chromosome 21 mapping and sequencing consortium consisting of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagamihara 228-8555, Japan,

\* e-mail: hattori@gsr.riken.go.jp

\* URL: http://hgp.gsc.riken.go.jp/

and

\* Institute of Molecular Biotechnology, Genome Analysis, \* Beutenbergstrasse 11, D-07745 Jena, Germany,

\* e-mail: gscj-submit@genome.imb-jena.de

\* URL: http://genome.imb-jena.de/

and

\* Keio University School of Medicine, Molecular Biology, \* Tokyo 160-8582, Japan,

\* e-mail: ashimizu@dbm-med.keio.ac.jp

\* URL: http://www.dmb.med.keio.ac.jp/

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e-mail: info.genome@gbf.de

\* URL: http://genome.gbf.de/

and

\* Max-Planck Institute for Molecular Genetics,

\* Innesstrasse 73, D-14195 Berlin, Germany,

\* e-mail: info-chr21@molgen.mpg.de

\* URL: http://chr21.rz-berlin.mpg.de/

AL163271: Submitted (10-Apr-2000).

Location/Qualifiers

1..340000

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21"

1..42441

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21"

## FEATURES

## source

## source

/map="21q22.1"  
/clone="KB5G11, 5' partial"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP000694"  
12310..112344  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="KB176G8"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP000695"  
62772..226523  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="KB1572B10"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP000696"  
184265..320526  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="KB594G10"  
/clone\_lib="Keio BAC library"  
/note="Accession No. AP000697"  
314337..>340000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="KUD8F6, 3' partial"  
/clone\_lib="KU21-D Cosmid library"  
/note="Accession No. AP000698"  
complement(748..1035)  
/note="Alu"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
1420..1699  
/note="AluSg"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement(1931..2227)  
/note="L1PA7"  
/rpt\_family="LINE/L1"  
/rpt\_type=DISPERSED  
2313..2509  
/note="AluJo"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
2737..2756  
/note="(TTTA)n"  
/rpt\_family="Simple\_repeat"  
/rpt\_type=TANDEM  
complement(2764..2880)  
/note="FLAM\_C"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement(2944..3064)  
/note="L1MA10"  
/rpt\_family="LINE/L1"  
/rpt\_type=DISPERSED  
complement(3350..3512)  
/note="AluJb"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
3699..3840  
/note="L1M4"  
/rpt\_family="LINE/L1"  
/rpt\_type=DISPERSED  
complement(4408..4556)



repeat_region	/note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(5287. .5353) /note="L1MB5" /rpt_family="LINE/L1" /rpt_type=DISPERSED 5354. .5561 /note="AluSg/x" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(5362. .5729) /note="L1MB5" /rpt_family="LINE/L1" /rpt_type=DISPERSED 5819. .6123 /note="AluJb" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(6129. .6417) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(6476. .6783) /note="L1MB5" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(7081. .7375) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 8000. .8284 /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 8285. .8313 /note="(CAAAA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(8576. .8642) /note="HAL1" /rpt_family="LINE/other" /rpt_type=DISPERSED complement(8643. .8938)	1612 ATCTCCAGAGCCACCGCGCAACACTGCTAGGCACAGCCCTGGTGGCCAGCTAGCAAGGTG 1671 1672 GGTCAAGTCTCTGCTGGGGAAGTGGGAGACTGCGCAGCGCGGAGCGGAGCGAGAGCC 1731 1732 ATGCGGTGGTGGCAGATGGAGACAGAACCTCACGCTTTGGCAAACTTGCCTCTTT 1791 1792 CTGCTTCTAAGTAGGCTTGTCTGCTTCTCAATGCAGGTCTCTCTGAGAGT 1851 1852 GAGAAATGCAGCTCTGCTGCTCGGCGACACTAGTAGCAGTATATAAGGGGCAAGGAAAA 1911 1912 CGAGTATCTGGCTTTCAGTAAATCTCTGGGCCACATTCACCAACCAAGGGGACAGTGA 1971 1972 TTTTCAAAACCCAGCTCCCATGTCTGAGAACACCCAGCTGCATTTCTTTGCAAGATTC 2031 2032 CTTTCCACTCCAAACCAAGAGTGAATATTTGAGACAAACGGGCTATTTGGCTATTTCCCAT 2091 2092 GCCAGTTTGGAAAGTGGGAAACTATGTGGAAATTTGTGGCTTTGGGCGACAGAAATGC 2151 2152 CACTCACCAACCCAGGCAAGAAACAAACCCCTCCAGGCTCAGTTTCTTCACTGTATA 2211 2212 AATGGGTGAAGCTGTATGTGCTACTCTCCAGGACACACACAGTAGGAGGACCTGCC 2271 2272 CTGTACATGCTAGTTTCAACAGAAAGAAATGGCTTTTCACTTCTCTCTGGTGGGCAAG 2331 2332 CAGATGCTCTGCGGAGATACCCAGCTCCCGAGCTCCCGAGCAGACTGACTCTCTTTGCT 2391 2392 CGCTGGACCAACCCAGCAGAGGTGGAAGTGGAAAGTGGGAGTGGAGTGGAGGACAT 2451 2452 GTATTCCCATTCGACCCGAGACTTAATCCAGCTCAGAGTGTAGACCCAGAGTGTGCAGAT 2511 2512 GCCTGCAGTGCATTAATAATGTGGTGAAGTGTAGATCAGGATTTATGTGCCCCAGGCGG 2571 2572 GCTCAGTGGCTCACACCTGTAAATCCAGCTTTGGGAGGCAAGTGGGCGGATCAGCT 2631 2632 GAGGTGAGAGTTTGGCAGACCTGCAACAGCTGAAACCCCTCTCCACTAAAAATA 2691 2692 CAAAAATTAGTTGGGCACTGGTGGAGCAGCTGTAAATCCAGCTACTTGGAGGCTGAGA 2751
repeat_region	Query Match Best Local Similarity 99.4%; Pred. No. 0; Matches 1599; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	1252 CCCACAGCAATACAGCTATTCCAAATGGACAACTGGAATGGCGGCGAGCTCGGAACT 1311 1312 GGAGAGCAGTCCCTGCGAGCGCTGCTCTCCAGAACTGCGAGCCCTCAGAAA 1371 1372 GCAGTGACTTCTGTACAGCGCATTCCTACAGCTGCGCTTCTCTACCATTTAGGACACT 1431 1432 TCCCTCTGGACTCTCACTTTTACAGAGCAAAAGCAATTTGCGGCGCAAGTTCGGGC 1491 1492 AGCCCAAGGATCCCTTGTGAGTGGCAGCCTTTTCTGAGCAAAATGCCAGCCAGCG 1551 1552 GTGAATGCCAGTGCAATATGCCAACCCCTAGTGCCTAGCAGCTCTCTCCAGCTAAAA 1611 1611 GTGAATGCCAGTGCAATATGCCAACCCCTAGTGCCTAGCAGCTCTCTCCAGCTAAAA 302809

QY	1612 ATCTCCAGAGCCACCGCGCAACACTGCTAGGCACAGCCCTGGTGGCCAGCTAGCAAGGTG 1671
Db	302810 ATCTCCAGAGCCACCGCGCAACACTGCTAGGCACAGCTGGTGGCCAGCTAGCAAGGTG 302869
QY	1672 GGTCAAGTCTCTGCTGGGGAAGTGGGAGACTGCGCAGCGCGGAGCGGAGAGCC 1731
Db	302870 GGTCAAGTCTCTGCTGGGGAAGTGGGAGACTGCGCAGCGCGGAGCGGAGAGCC 302929
QY	1732 ATGCGGTGGTGGCAGATGGAGACAGAACCTCACGCTTTGGCAAACTTGCCTCTTT 1791
Db	302930 ATGCGGTGGTGGCAGATGGAGACAGAACCTCACGCTTTGGCAAACTTGCCTCTTT 302989
QY	1792 CTGCTTCTAAGTAGGCTTGTCTGCTTCTCAATGCAGGTCTCTCTGAGAGT 1851
Db	303049 CTGCTTCTAAGTAGGCTTGTCTGCTTCTCAATGCAGGTCTCTCTGAGAGT 303049
QY	1852 GAGAAATGCAGCTCTGCTGCTCGGCGACACTAGTAGCAGTATATAAGGGGCAAGGAAAA 1911
Db	303050 GAGAAATGCAGCTCTGCTGCTCGGCGACACTAGTAGCAGTATATAAGGGGCAAGGAAAA 303109
QY	1912 CGAGTATCTGGCTTTCAGTAAATCTCTGGGCCACATTCACCAACCAAGGGGACAGTGA 1971
Db	303110 CGAGTATCTGGCTTTCAGTAAATCTCTGGGCCACATTCACCAACCAAGGGGACAGTGA 303169
QY	1972 TTTTCAAAACCCAGCTCCCATGTCTGAGAACACCCAGCTGCATTTCTTTGCAAGATTC 2031
Db	303170 TTTTCAAAACCCAGCTCCCATGTCTGAGAACACCCAGCTGCATTTCTTTGCAAGATTC 303229
QY	2032 CTTTCCACTCCAAACCAAGAGTGAATATTTGAGACAAACGGGCTATTTGGCTATTTCCCAT 2091
Db	303230 CTTTCCACTCCAAACCAAGAGTGAATATTTGAGACAAACGGGCTATTTGGCTATTTCCCAT 303289
QY	2092 GCCAGTTTGGAAAGTGGGAAACTATGTGGAAATTTGTGGCTTTGGGCGACAGAAATGC 2151
Db	303290 GCCAGTTTGGAAAGTGGGAAACTATGTGGAAATTTGTGGCTTTGGGCGACAGAAATGC 303349
QY	2152 CACTCACCAACCCAGGCAAGAAACAAACCCCTCCAGGCTCAGTTTCTTCACTGTATA 2211
Db	303350 CACTCACCAACCCAGGCAAGAAACAAACCCCTCCAGGCTCAGTTTCTTCACTGTATA 303409
QY	2212 AATGGGTGAAGCTGTATGTGCTACTCTCCAGGACACACACAGTAGGAGGACCTGCC 2271
Db	303410 AATGGGTGAAGCTGTATGTGCTACTCTCCAGGACACACACAGTAGGAGGACCTGCC 303469
QY	2272 CTGTACATGCTAGTTTCAACAGAAAGAAATGGCTTTTCACTTCTCTCTGGTGGGCAAG 2331
Db	303470 CTGTACATGCTAGTTTCAACAGAAAGAAATGGCTTTTCACTTCTCTCTGGTGGGCAAG 303529
QY	2332 CAGATGCTCTGCGGAGATACCCAGCTCCCGAGCTCCCGAGCAGACTGACTCTCTTTGCT 2391
Db	303530 CAGATGCTCTGCGGAGATACCCAGCTCCCGAGCTCCCGAGCAGACTGACTCTCTTTGCT 303589
QY	2392 CGCTGGACCAACCCAGCAGAGGTGGAAGTGGAAAGTGGGAGTGGAGTGGAGGACAT 2451
Db	303590 CGCTGGACCAACCCAGCAGAGGTGGAAGTGGAAAGTGGGAGTGGAGTGGAGGACAT 303649
QY	2452 GTATTCCCATTCGACCCGAGACTTAATCCAGCTCAGAGTGTAGACCCAGAGTGTGCAGAT 2511
Db	303650 GTATTCCCATTCGACCCGAGACTTAATCCAGCTCAGAGTGTAGACCCAGAGTGTGCAGAT 303709
QY	2512 GCCTGCAGTGCATTAATAATGTGGTGAAGTGTAGATCAGGATTTATGTGCCCCAGGCGG 2571
Db	303710 GCCTGCAGTGCATTAATAATGTGGTGAAGTGTAGATCAGGATTTATGTGCCCCAGGCGG 303769
QY	2572 GCTCAGTGGCTCACACCTGTAAATCCAGCTTTGGGAGGCAAGTGGGCGGATCAGCT 2631
Db	303770 GCTCAGTGGCTCACACCTGTAAATCCAGCTTTGGGAGGCAAGTGGGCGGATCAGCT 303829
QY	2632 GAGGTGAGAGTTTGGCAGACCTGCAACAGCTGAAACCCCTCTCCACTAAAAATA 2691
Db	303830 GAGGTGAGAGTTTGGCAGACCTGCAACAGCTGAAACCCCTCTCCACTAAAAATA 303889
QY	2692 CAAAAATTAGTTGGGCACTGGTGGAGCAGCTGTAAATCCAGCTACTTGGAGGCTGAGA 2751



||||| 303890 CAAAAATTAGTTGGGCATGGTGGTGGACACCTGTATCCAGCTACTCTGGAGGCTGAGA 303949  
QY 2752 TAGGAGGATCACTTGAACCCGGGAGGTGGAGGTTCAGTGTAGATCAAGATCAATCACTGTC 2811  
Db 303950 TAGGAGGATCACTTGAACCCGGGAGGTGGAGGTTCAGTGTAGATCAATCACTGTC 304009  
QY 2812 ACTCCAGCCTGGGTACAGAGTGGAGCTGTCTCAAAAAA 2859  
Db 304010 ACTCCAGCCTGGGTACAGAGTGGAGCTGTCTCAAAAAA 304057  
  
RESULT 5  
AB003185 3365 bp DNA linear PRI 14-APR-2000  
LOCUS Homo sapiens DNA for SIM2, SIM2s, partial cds.  
ACCESSION AB003185  
VERSION AB003185.1 GI:2257468  
KEYWORDS alternative splicing; SIM2; SIM2s; SIM2 protein.  
SOURCE Homo sapiens pre-pro-B cell cell\_line: FLEB14-14 DNA, clone\_lib: BAC library clone: KB594G10.  
  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
  
REFERENCE 1 (sites)  
AUTHORS Chast.R., Scott.H.S., Chen.H., Kudoh.J., Rossier.C., Minoshima.S.,  
Wang.Y., Shimizu.N. and Antonarakis.S.E.  
TITLE Cloning of two human homologs of the Drosophila single-minded gene  
SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome  
chromosomal region  
JOURNAL Genome Res. 7 (6), 615-624 (1997)  
MEDLINE 97343329  
AUTHORS Chast.R., Scott.H.S., Chen.H., Kudoh.J., Rossier.C., Minoshima.S.,  
Wang.Y., Shimizu.N. and Antonarakis.S.E.  
TITLE Direct Submission  
JOURNAL Submitted (18-APR-1997) Nobuyoshi Shimizu, Kelo University School  
of Medicine, Department of Molecular Biology; 35 Shinanomachi,  
Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@db.med.keio.ac.jp,  
Tel:03-3351-2370, Fax:03-3351-2370)  
FEATURES Location/Qualifiers  
source 1..3365  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1-22.2"  
/clone="KB594G10"  
/cell\_line="FLEB14-14"  
/cell\_type="pre-pro-B cell"  
/clone\_lib="BAC library"  
join(<1..409,2938..3365)  
/note="alternative splicing"  
/codon\_start=1  
/product="SIM2"  
/protein\_id="BAA21489.1"  
/db\_xref="GI:2257469"  
/translation="OYSSFQMDKLECGQGNWRASPPASAAAPPELOPHSSDLLYT  
PSYSLPFSYHYGHFPLDSHVSSKPKMLPAKPGQSPCEVARFFLTSLPASGECW  
HYANPLVPSSSSPAKNPPEPPANTARHSLVPSYEAAPAAVRFEGEDTAPFPSPGCHY  
REPLALGPAKAAQARDGALALAAPECCAPPTPEAPGAPQLPFVLLNHYRLA  
RRGLGGAAPAAASGLACAPGPGPEATGALRLRHPSPAATSPPGAPLPHYLGASVLIIN  
Gr"  
<1..546  
/note="alternative splicing"  
/codon\_start=1  
/product="SIM2s"  
/protein\_id="BAA21490.1"  
/db\_xref="GI:2257470"  
/translation="OYSSFQMDKLECGQGNWRASPPASAAAPPELOPHSSDLLYT  
PSYSLPFSYHYGHFPLDSHVSSKPKMLPAKPGQSPCEVARFFLTSLPASGECW  
HYANPLVPSSSSPAKNPPEPPANTARHSLVPSYEGSGLLVGVGGLRTAGSRSSHGG  
GWQWETPSRFQTCPLSASK"  
1..409  
  
exon

intron /number=10  
410..2937  
exon /number=10  
2938..3365  
/number=11  
BASE COUNT 806 a 966 c 915 g 678 t  
ORIGIN  
  
Query Match 55.6%; Score 1590.4; DB 9; Length 3365;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1594; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1260 CAATACAGCTCATTCCTCAAAATGGACAAATGGGAATCGGCGACGCTCGGAACCTGGAGAGCC 1319  
Db 1 CAATACAGCTCGTTCCTCAAAATGGACAAATGGGAATCGGCGACGCTCGGAACCTGGAGAGCC 60  
QY 1320 AGTCCCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCAGCCCTCACTCAAGAAAGCAGTGAC 1379  
Db 61 AGTCCCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCAGCCCTCACTCAAGAAAGCAGTGAC 120  
QY 1380 CTTCTGTACACGCCATCTTACAGCTGCGCTTCTCTCTACCATTTATGGACACTTCCTCTG 1439  
Db 121 CTTCTGTACACGCCATCTTACAGCTGCGCTTCTCTCTACCATTTATGGACACTTCCTCTG 180  
QY 1440 GACTCTCACTTCTTTCAGCAGCAAAAGCAATGTTGCCGGCGCAAGTTCCGGCAGCCCA 1499  
Db 181 GACTCTCACTTCTTTCAGCAGCAAAAGCAATGTTGCCGGCGCAAGTTCCGGCAGCCCA 240  
QY 1500 GGATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCACTGCGAGCGGCGGTAATGC 1559  
Db 241 GGATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCACTGCGAGCGGCGGTAATGC 300  
QY 1560 CAGTGGCATTTATGCCAACCCCTTAGTGCTAGCAGCTGCTCTCAGCTCAAAATCTCCCA 1619  
Db 301 CAGTGGCATTTATGCCAACCCCTTAGTGCTAGCAGCTGCTCTCAGCTCAAAATCTCCCA 360  
QY 1620 GAGCCACCGGCAACACTGCTAGGACAGCCTGCTGCAAGCTACGAGGTGGGTCAAGT 1679  
Db 361 GAGCCACCGGCAACACTGCTAGGACAGCCTGCTGCAAGCTACGAGGTGGGTCAAGT 420  
QY 1680 CTGCTCTGTTGGGAAAGTGGGAGGACTGCGACAGCGCGGCGCAAGCAAGCCATGGCGT 1739  
Db 421 CTGCTCTGTTGGGAAAGTGGGAGGACTGCGACAGCGCGGCGCAAGCAAGCCATGGCGT 480  
QY 1740 GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCCTCTTTCTGCTTCT 1799  
Db 481 GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCCTCTTTCTGCTTCT 540  
QY 1800 AAGTAGGGCTTGTGCTGCTTCTTGTCTCAATGCAGGTGCTCTCGAGAGTGAGAAATG 1859  
Db 541 AAGTAGGGCTTGTGCTGCTTCTTGTCTCAATGCAGGTGCTCTCGAGAGTGAGAAATG 600  
QY 1860 GCATCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGCGCAAGGAAACCGAGTAT 1919  
Db 601 GCATCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGCGCAAGGAAACCGAGTAT 560  
QY 1920 CTGSCCTTCACGTAATCTCTGGCCACATTCACCAACCAAGGGGGAGTGTATTTTCAA 1979  
Db 661 CTGSCCTTCACGTAATCTCTGGCCACATTCACCAACCAAGGGGGAGTGTATTTTCAA 720  
QY 1980 ACCAGCTCCCATGCTGTGAGAACACCCAGCTGCTTTTTCGAAAGATTCCTTTCCAC 2039  
Db 721 ACCAGCTCCCATGCTGTGAGAACACCCAGCTGCTTTTTCGAAAGATTCCTTTCCAC 780  
QY 2040 TCCAACCCAGAGTGAATATTTGAGACAAACGGGCTATTGGCTATTTTCCCATGCCAGTTT 2099  
Db 781 TCCAACCCAGAGTGAATATTTGAGACAAACGGGCTATTGGCTATTTTCCCATGCCAGTTT 840  
QY 2100 TGAAGTGGGGAAAACTAGTGGAAATTTGTGGCTTGGGACAGAAATGCCACTCACC 2159  
Db 841 TGAAGTGGGGAAAACTAGTGGAAATTTGTGGCTTGGGACAGAAATGCCACTCACC 900  
QY 2160 AATCCAGGGGAAAGAACACAAACCCCTCCAGGCTCAGTTTCTTCACTGTAAATGGGCT 2219



117760	TCCCTCTGGACTCTCACGCTCTTTCAGCAGCAAAAAGCCAATGTTGCGGCCCAAGTTTCGGCG	117819	
QY	1492	AGCCCCAAGGATCCCTTGTGAGGTGGCAGCGCTTTTTCCTGAGCACAATGCGCAGCCAGCG	1551
Db	117820	AGCCCCAAGGATCCCTTGTGAGGTGGCAGCGCTTTTTCCTGAGCACAATGCGCAGCCAGCG	117879
QY	1552	GTGAATGCCAGTGGCAATATGCGAACCCCTTGTGCTAGTGCCTAGCAGCTCGTCTCCAGCTAAAA	1611
Db	117880	GTGAATGCCAGTGGCAATATGCGAACCCCTTGTGCTAGTGCCTAGCAGCTCGTCTCCAGCTAAAA	117939
QY	1612	ATCCTCCAGAGCCACCGGCAACACTGCTTAGCAGCAGCCTGTGTGCCAAGCTACGAAAGTG	1671
Db	117940	ATCCTCCAGAGCCACCGGCAACACTGCTTAGCAGCAGCCTGTGTGCCAAGCTACGAAAGTG	117999
QY	1672	GGTGAGTCTGCTGTGGGAAAGTGGAGGACTGCGCAGCGCGGGAGCGGAGCAGCC	1731
Db	118000	GGTGAGTCTGCTGTGGGAAAGTGGAGGACTGCGCAGCGCGGGAGCGGAGCAGCC	118059
QY	1732	ATGGCGGTGGTGGAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCTCTTT	1791
Db	118060	ATGGCGGTGGTGGAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCTCTTT	118119
QY	1792	CTGCTTCTAAGTAGGCTTGTCTTCTTCTTCTCAATGCAAGTGCCTTCGAGAGT	1851
Db	118120	CTGCTTCTAAGTAGGCTTGTCTTCTTCTTCTCAATGCAAGTGCCTTCGAGAGT	118179
QY	1852	GAGAAATGCACTGCTGCTCGGGGACACTAGTGACAGTATAAAGGCAAGGAAAA	1911
Db	118180	GAGAAATGCACTGCTGCTCGGGGACACTAGTGACAGTATAAAGGCAAGGAAAA	118239
QY	1912	CGAGTATCTGCGCTTTCAGCTAAATCTCGGCCCACTTCCACCAAGGGGACAGTGA	1971
Db	118240	CGAGTATCTGCGCTTTCAGCTAAATCTCGGCCCACTTCCACCAAGGGGACAGTGA	118299
QY	1972	TTTTCAAAACAGCTCCCATGTGCTGAGAACCCAGCTGCATTTCTTTGCAAGATTC	2031
Db	118300	TTTTCAAAACAGCTCCCATGTGCTGAGAACCCAGCTGCATTTCTTTGCAAGATTC	118359
QY	2032	CTTTCCACTTCAACAGAGTGAATATTTGAGACAACGGCCTATTGGCTATTTTCCCAT	2091
Db	118360	CTTTCCACTTCAACAGAGTGAATATTTGAGACAACGGCCTATTGGCTATTTTCCCAT	118419
QY	2092	GCCAGTTTGGAGTGGGAAACATATGTTGGAAATTTGGGCTTGGGACAGAAATGC	2151
Db	118420	GCCAGTTTGGAGTGGGAAACATATGTTGGAAATTTGGGCTTGGGACAGAAATGC	118479
QY	2152	CACACCAACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTCACCTGTAA	2211
Db	118480	CACACCAACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTCACCTGTAA	118539
QY	2212	AATGGGTGAAGCTGTGATGTGCTACTCCAGGACACAGACACAGTAGGAGCTGCG	2271
Db	118540	AATGGGTGAAGCTGTGATGTGCTACTCCAGGACACAGACACAGTAGGAGCTGCG	118599
QY	2272	CTGTACATGCTAGTTCAACAGAAAGGAATGGCTTTTCACTTCTCTGTGGCAGGCAAG	2331
Db	118600	CTGTACATGCTAGTTCAACAGAAAGGAATGGCTTTTCACTTCTCTGTGGCAGGCAAG	118659
QY	2332	CAGATGTCTCTCGGAGATACCGCAGCTCCCGAGGACGACAGCTCCTGTGTTGCT	2391
Db	118660	CAGATGTCTCTCGGAGATACCGCAGCTCCCGAGGACGACAGCTCCTGTGTTGCT	118719
QY	2392	CGTGGACCAACCCCGAGGAGAGTGGAGGTGGGAACAGAGGTTTACGTCGAGGACAT	2451
Db	118720	CGTGGACCAACCCCGAGGAGAGTGGAGGTGGGAACAGAGGTTTACGTCGAGGACAT	118779
QY	2452	GTATTTCCCATTCACCGAGACCTAATGCGGCTCAGAGTGTAGACCGAGATGGTCAGAT	2511
Db	118780	GTATTTCCCATTCACCGAGACCTAATGCGGCTCAGAGTGTAGACCGAGATGGTCAGAT	118839
QY	2512	GCCTGAGTGCCTATTAATGTTGGTGAAGTGACATCAGATTTATGTGCCCGAGCGCG	2571
Db	118840	GTCTGAGTGCCTATTAATGTTGGTGAAGTGACATCAGATTTATGTGCCCGAGCGCG	118899

BASE COUNT		345 a	446 c	376 g	298 t
ORIGIN					
Query Match					
Best Local Similarity 100.0%; Score 1389; DB 9; Length 1465;					
Matches 1389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	281	GTGGGGACAGCCGAGCCGCGCGCCCTGGAGCGGTGCGCAAGAGAGCTGGATCGCA	340		
DB	1	GTGGGGACAGCCGAGCCGCGCGCCCTGGAGCGGTGCGCAAGAGAGCTGGATCGCA	60		
QY	341	CTTGCTGCAGACTTTGGATGATTTGTTTGTGTAGTACTCTGATGGCAAAATCATGTA	400		
DB	61	CTTGCTGCAGACTTTGGATGATTTGTTTGTGTAGTACTCTGATGGCAAAATCATGTA	120		
QY	401	TATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGAGCTCAGGGCAACAG	460		
DB	121	TATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGAGCTCAGGGCAACAG	180		
QY	461	TATTTATGAATACATCCATCTCTGACACGATGAGATGACCGCTCTCTCAGGGCCCA	520		
DB	181	TATTTATGAATACATCCATCTCTGACACGATGAGATGACCGCTCTCTCAGGGCCCA	240		
QY	521	CCAGCGCTGTCACCAACCGCTGCTCCAAAGATATGAGATGAGAGTCTTCTTCTCG	580		
DB	241	CCAGCGCTGTCACCAACCGCTGCTCCAAAGATATGAGATGAGAGTCTTCTTCTCG	300		
QY	581	AATGAATGTGTTGGCGAAAGAACCGGGCTGACCTGCAGCGGATACAAAGTCTAT	640		
DB	301	AATGAATGTGTTGGCGAAAGAACCGGGCTGACCTGCAGCGGATACAAAGTCTAT	360		
QY	641	CCACTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGACATGTCCTGTGACGACTC	700		
DB	361	CCACTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGACATGTCCTGTGACGACTC	420		
QY	701	CTGCTACCAAGATGTGGGCTGTGGCGTGGCCAGTCTGCTGCCACCGAGTGCATCAC	760		
DB	421	CTGCTACCAAGATGTGGGCTGTGGCGTGGCCAGTCTGCTGCCACCGAGTGCATCAC	480		
QY	761	CGAGATCAAGCTGTACAGTAACATGTTTCATGTTCAGGGCCAGCGCTTGACCTGAACCTGAT	820		
DB	481	CGAGATCAAGCTGTACAGTAACATGTTTCATGTTCAGGGCCAGCGCTTGACCTGAACCTGAT	540		
QY	821	ATTCTCGGATTCAGGGTGACCGAGTACGGCTTACGAGCCGAGGAGCTGATCGAGAA	880		
DB	541	ATTCTCGGATTCAGGGTGACCGAGTACGGCTTACGAGCCGAGGAGCTGATCGAGAA	600		
QY	881	GACCTATACCATCAGTGCACCGCTGCGACGCTGTTCCACCTCCGCTACGCGACACCACT	940		
DB	601	GACCTATACCATCAGTGCACCGCTGCGAGCGTGTCCACCTCCGCTACGCGACACCACT	660		
QY	941	CTGTGTGGTGAAGGGCCAGGTACACCAAGTACTACCGGCTGCTGTCGAAGGGGGCGG	1000		
DB	661	CTGTGTGGTGAAGGGCCAGGTACACCAAGTACTACCGGCTGCTGTCGAAGGGGGCGG	720		
QY	1001	CTGGGTGTGGTGCAGAGTACGCCACCGTGTGTGCACACAGCGCTGCTCCGGGCCCA	1060		
DB	721	CTGGGTGTGGTGCAGAGTACGCCACCGTGTGTGCACACAGCGCTGCTCCGGGCCCA	780		
QY	1061	CTGCATCGTGAAGTCAATATGTACTACGGAGATTGAATACAGGAATTCAGCTGTC	1120		
DB	781	CTGCATCGTGAAGTCAATATGTACTACGGAGATTGAATACAGGAATTCAGCTGTC	840		
QY	1121	CCTGGACAGGTGTCCACTGCGCAAGTCCAGGACTCCTGGAGGACCGCTTGTCTACCTC	1180		
DB	841	CCTGGACAGGTGTCCACTGCGCAAGTCCAGGACTCCTGGAGGACCGCTTGTCTACCTC	900		
QY	1181	ACAAGAACTAGGAATTTAGTGAACCCCAAAATPACCAAGATGAAGCAAGCTGGAAC	1240		
DB	901	ACAAGAACTAGGAATTTAGTGAACCCCAAAATPACCAAGATGAAGCAAGCTGGAAC	960		

QY	1241	AAACCCCTTACCCCCACAGCAATACAGCTCATTCCTCAATGACAACTGGATCGGCCA	1300		
DB	961	AAACCCCTTACCCCCACAGCAATACAGCTCATTCCTCAATGACAACTGGATCGGCCA	1020		
QY	1301	GCTCGGAAACTGGAGAGCCAGTCCCCCTGCAAGCGCTGCTCTCCAGAACTGCGGCC	1360		
DB	1021	GCTCGGAAACTGGAGAGCCAGTCCCCCTGCAAGCGCTGCTCTCCAGAACTGCGGCC	1080		
QY	1361	CCACTCAGAAAGCACTGACCTTCTGTACACGCCATCTTACAGCCTGCGCTTCTCTTACCA	1420		
DB	1081	CCACTCAGAAAGCACTGACCTTCTGTACACGCCATCTTACAGCCTGCGCTTCTCTTACCA	1140		
QY	1421	TTATGACACTTCCCTCTGGACTCTCACTTCTTACAGCAAGCAAAAGCAATGTTGCCGCGC	1480		
DB	1141	TTATGACACTTCCCTCTGGACTCTCACTTCTTACAGCAAGCAAAAGCAATGTTGCCGCGC	1200		
QY	1481	CAAGTTCGGGCGACCCCAAGGATCCCTTGTGAGTGGCAGCCTTTTCTCTGAGCACAAT	1540		
DB	1201	CAAGTTCGGGCGACCCCAAGGATCCCTTGTGAGTGGCAGCCTTTTCTCTGAGCACAAT	1260		
QY	1541	GCCAGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTGTC	1600		
DB	1261	GCCAGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTGTC	1320		
QY	1601	TCCAGCTAAATCTCTCCAGAGCCAGCCGCGGACACTGCTAGCAGCCTGCTGTCGAAG	1660		
DB	1321	TCCAGCTAAATCTCTCCAGAGCCAGCCGCGGACACTGCTAGCAGCCTGCTGTCGAAG	1380		
QY	1661	CTACGAAGG 1669			
DB	1381	CTACGAAGG 1389			
RESULT 8					
LOCUS	E14065	3614 bp	DNA	linear	PAT 28-JUL-1999
DEFINITION	cDNA encoding SIM protein.				
ACCESSION	E14065				
VERSION	E14065.1	GI:5708748			
KEYWORDS	JP 1997268199-A/7.				
SOURCE	Mus sp.				
ORGANISM	Mus sp.				
REFERENCE	1 (bases 1 to 3614)				
AUTHORS	Shimizu, N.				
TITLE	HUMAN OR MOUSE SIM GENE				
JOURNAL	Patent: JP 1997268199-A 7 14-OCT-1997;				
COMMENT	FUJISAWA PHARMACEUT CO LTD				
	OS Mus sp. (mouse)				
	PN JP 1997268199-A/7				
	PD 14-OCT-1997				
	PF 28-MAR-1996 JP 1996074261				
	PR 31-MAR-1995 JP 95P 76362, 29-JAN-1996 JP 96P 12952 PI				
	SHIMIZU NOBUYOSHI				
	PC C07K14/47, C12N1/21, C12N15/09, C12Q1/68, C12P21/02, (C12N1/21, PC				
	C12R1:19);				
	CC strandedness: Double;				
	CC topology: Linear;				
	CC hypothetical: No;				
	CC anti-sense: No;				
	FH Key				
	FH Location/Qualifiers				
	FT source				
	FT 1..3614				
	FT /organism='Mus sp.'				
	FT /tissue_type='embryo'				
	FT 1..503				
	FT 5'UTR				
	FT CDS				
	FT 504..2474				
	FT /product='SIM protein'				
	FT 2475..3614.				
	FT 3'UTR				
	FT Location/Qualifiers				
	FT 1..3614				
FEATURES	source				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Query Match		/organism="Mus sp."		833 a 1110 c 952 g 717 t 2 others	
Best Local Similarity		46.3%; Score 1323.4; DB 6; Length 3614;			
Matches 1451; Conservative		Pred. No. 7.4e-303;			
		0; Mismatches 191; Indels 3; Gaps 1;			
QY	27	CGGGGAGGTGGGCTCCCGGGCTTGAGCAGCGCGGGTCTAATATGCCGGAGCGGA	86		
Db	438	CAGCCACAGGGAGCTCCGGGGCTTGAGCGCAGCAGGCTAATATGCCAGAGCGGA	497		
QY	87	GGCGGATGAAGGAGAAGTCCAAAGATCGCGCAAGACAGAGGAGGAGAAATGGC	146		
Db	498	GGCGGATGAAGGAGAAGTCCAAAGATCGCGCAAGACAGAGGAGGAGAAATGGC	557		
QY	147	GAGTTTACGAGCTTGCCAAAGTGTCCCGTGGCGTCCGACATCACTTCCAGCTGGAC	206		
Db	558	GAGTTTACGAGCTTGCCAAAGTGTCCCGTGGCGTCCGACATCACTTCCAGCTGGAC	617		
QY	207	AAAGCGTCCATCATCCGCTCACCACGAGCTACCTGAAGATGCCGGCGTCTTCCCGAA	266		
Db	618	AAAGCGTCCATCATCCGCTCACCACGAGCTACCTGAAGATGCCGGCGTCTTCCCGAA	677		
QY	267	GGTTTAGGAGAGCGGTGGGACAGCGGAGCGCGCGCCCTGGAGCGCGTGGCCAAAG	326		
Db	678	GGTTTAGGAGAGCGGTGGGACAGCGGAGCGCGCGCCCTGGAGCGCGTGGCCAAAG	737		
QY	327	GAGCTGGGATCGCACTTGTCCAGACTTGGATGGATTTGTTTGTGGTGGATCTGAT	386		
Db	738	GAGCTGGGATCGCACTTGTCCAGACTTGGATGGATTTGTTTGTGGTGGATCTGAT	797		
QY	387	GGCAAAATCATATATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGGAG	446		
Db	798	GGCAAAATCATATATATCTGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGGAG	857		
QY	447	CTCAGGGAACAGTATTATGATATCATCCATCTCTTGACCAAGATGAGATGACCGT	506		
Db	858	CTCAGGGAACAGTATTATGATATCATCCATCTCTTGACCAAGATGAGATGACCGT	917		
QY	507	GTCTCTACGGCCACCAAGCGCTGCACACCACTCTGTCACCAAGATGAGATGAGAGG	566		
Db	918	GTCTCTACGGCCACCAAGCGCTGCACACCACTCTGTCACCAAGATGAGATGAGAGG	977		
QY	567	TCGTTCTTTCTGGAATGATGTCTTGGGAAAGGAAAGCGCGGCTGACCTCGACG	626		
Db	978	TCGTTCTTTCTGGAATGATGTCTTGGGAAAGGAAAGCGCGGCTGACCTCGACG	1037		
QY	627	GGATACAGGTGATCCAGTGCAGTGTCTGTAAGATCAGGAGTATATGCTGGACATG	686		
Db	1038	GGATACAGGTGATCCAGTGCAGTGTCTGTAAGATCAGGAGTATATGCTGGACATG	1097		
QY	687	TCCTGTACGACTCTCTACCAAGATGTGGGGCTGGTGGCGGTGGCGGCTGCGCA	746		
Db	1098	TCCTGTACGACTCTCTACCAAGATGTGGGGCTGGTGGCGGTGGCGGCTGCGCA	1157		
QY	747	CCAGTGGCATCACCAGATCAAGCTGTACATTAATGTTTATGCTTTCAGGCGGCTT	806		
Db	1158	CCAGTGGCATCACCAGATCAAGCTGTACATTAATGTTTATGCTTTCAGGCGGCTT	1217		
QY	807	GACCTGAAGCTGATATTTCTGATTCAGGCTGACCGAGGTACCGGTTACGAGCCGAG	866		
Db	1218	GACCTGAAGCTGATATTTCTGATTCAGGCTGACCGAGGTACCGGTTACGAGCCGAG	1277		
QY	867	GACCTGATCGAAGAGACCTTATACCATCAGCTGACGCTGCGAGCTTCCACCTCCG	926		
Db	1278	GACCTGATCGAAGAGACCTTATACCATCAGCTGACGCTGCGAGCTTCCACCTCCG	1337		
QY	927	TACGACACACCTCTCTGTTGGTGAAGGCGGAGTCAACCAAGTACTACCGCTGCTG	986		
Db	1338	TATGACACACCTCTCTGTTGGTGAAGGCGGAGTCAACCAAGTACTACCGCTGCTG	1397		

QY	987	TCCAAGCGGGCGCTGGGTGCTGGGTGCAGAGCTACGCCACCGTGTGTGCACACAGCCGC	1046
Db	1398	TCCAAGCTGGCGGCTGGGTGCTGGGTGCAGAGCTACGCCACCGTGTGTGCACACAGCCGC	1457
QY	1047	TCGTCCGGCGCCACATCGTGAAGTCAATATGATCAGCGAGATTCGAATACAG	1106
Db	1458	TCGTCCGGCGCCACATCGTGAAGTCAATATGATCAGCGAGATTCGAATACAG	1517
QY	1107	GAACTTCAGCTGCTCCCTGGAGAGGTGTCACATGCCAAGTCCAGAGCTCTCTGGAGACC	1166
Db	1518	GAACTTCAGCTGCTCCCTGGAGAGGTGTCACATGCCAAGTCCAGAGCTCTCTGGAGACC	1577
QY	1167	GCCTTGTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	1226
Db	1578	GCCTTGTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	1637
QY	1227	ACAAAGCTCAGAACAAACCTTACCCCCACAGCAATACAGCTCTTCCAAATGGACAA	1286
Db	1638	ACAAAGCTCAGAACAAACCTTACCCCCACAGCAATACAGCTCTTCCAAATGGACAA	1697
QY	1287	CTGGAATCGGCGCAGCTCGGAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCT	1346
Db	1698	CTGGAATCGGCGCAGCTCGGAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCT	1757
QY	1347	CCAGAACTCCAGCCCGCCTCAGAAAGCAGTACACCTTCTGTACAGCCCATCTACAGCTG	1406
Db	1758	CCAGAACTCCAGCCCGCCTCAGAAAGCAGTACACCTTCTGTACAGCCCATCTACAGCTG	1817
QY	1407	CCCTTCTCTACCATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTCCAGCAGCAAAAG	1466
Db	1818	CCCTTCTCTACCATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTCCAGCAGCAAAAG	1877
QY	1467	CCAAATGTTCCGGCGCAAGTTCGGGCGAGCCCAAGATCCCTTGTGAGGTGGCAGCTTT	1526
Db	1878	CCAAATGTTCCGGCGCAAGTTCGGGCGAGCCCAAGATCCCTTGTGAGGTGGCAGCTTT	1937
QY	1527	TTCCTGAGCACAATGCCAGCGGTGAATGCCAGTGGCATATATGCCAACCCCTTAGT	1586
Db	1938	TTCCTGAGCACAATGCCAGCGGTGAATGCCAGTGGCATATATGCCAACCCCTTAGT	1997
QY	1587	CCTAGCAGCTGCTTCCAGCTTAAATCTCCACAGC---CACGGCGCAACTGCTAGG	1643
Db	1998	CCTAGCAGCTGCTTCCAGCTTAAATCTCCACAGC---CACGGCGCAACTGCTAGG	2057
QY	1644	CACAGCTGCTGCGCAAGCTACGAAG	1668
Db	2058	CATGCTCTGTGCGCAAACTATGAAG	2082

## RESULT 9

LOCUS	D64135	3614 bp	mRNA	linear	ROD 13-FEB-1999
DEFINITION	Mus musculus Sim2 mRNA, complete cds.				
ACCESSION	D64135				
VERSION	D64135.1				
KEYWORDS	Sim2.				
SOURCE	Mus musculus (strain:ICR x Swiss Webster) whole embryo cDNA to mRNA, clone_lib:5'-stretch cDNA library (CLONTECH).				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	1 (sites)				
TITLE	Yamaki, A., Noda, S., Kudoh, J., Shindoh, N., Maeda, H., Minoshima, S., Kawasaki, K., Shimizu, Y. and Shimizu, N.				
JOURNAL	The mammalian single-minded (SIM) gene: mouse cDNA structure and				
MEDLINE	Genomics 35 (1), 136-143 (1996)				
REFERENCE	96299750				
AUTHORS	2 (bases 1 to 3614)				
TITLE	Shimizu, N.				
JOURNAL	Direct Submission				
	Submitted (13-SEP-1995) Nobuyoshi Shimizu, Keio University School				
	of Medicine, Department of Molecular Biology; 35 Shinanomachi,				

Shinjuku-ku, Tokyo 160-8582, Japan (E-mail:shimizmedb.med.keio.ac.jp, Tel:81-3-3351-2370(ex.2720), Fax:81-3-3351-2370)		QY		567	TCGTTCTTTCTCGAATGAAATGTCTTGGCGAAAGAAACGCGGCGCTGACCTGACG 626
FEATURES		Db		978	TCCTTTCTTTCTCGAATGAAATGTCTTGGCGAAAGAAACGCGGCGCTGACATGAGT 1037
source		QY		627	GGATACAAGGTCAATCCACTGCAGTGGCTACTTTGAAGATCAGGACAGTATATCTGGACATG 686
Location/Qualifiers		Db		1038	GGATACAAGGTCAATCCACTGCAGTGGCTACTTTGAAGATCAGGACAGTATATCTGGACATG 1097
/organism="Mus musculus"		QY		687	TCCTTGTAGGACTCTCTGCTACAGATTTGGGGCTTGGTGGCGTGGGCGAGTTCGCTGCCA 746
/strain="ICR x Swiss Webster"		Db		1098	TCCTTGTAGGACTCTCTGCTACAGATTTGGGGCTTGGTGGCGTGGGCGAGTTCGCTGCCA 1157
/db_xref="taxon:10090"		QY		747	CCCAGTGCATCACGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTCAGGGCCAGCCTT 806
/tissue_type="whole embryo"		Db		1158	CCCAGTGCATCACGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTCAGGGCCAGCCTT 1217
/clone_lib="5'-stretch cDNA library (CLONTECH)"		QY		807	GACCTGAAGCTCATATTCCTGATTTCCAGGTTCCAGGTTGACGAGTTCACGAGCCGAG 866
1..3614		Db		1218	GACCTGAAGCTCATATTCCTGATTTCCAGGTTCCAGGTTGACGAGTTCACGAGCCGAG 1277
/gene="Sim2"		QY		867	GACCTGCAGAGAACCTTATACCATCATGTCATGCTGTGACACCTTCCACCTCCGCA 926
504..2477		Db		1278	GACCTGCAGAGAACCTTATACCATCATGTCATGCTGTGACACCTTCCACCTCCGCA 1337
/standard_name="single-minded gene"		QY		927	TAGCACACACCTCTCTGTTGGTGAAGGCGCAGGTCACCAACCAAGTACTACCGGTGCTG 986
/codon_start=1		Db		1338	TATGCACACACCTCTCTGTTGGTGAAGGCGCAGGTCACCAACCAAGTACTACCGGTGCTG 1397
/protein_id="BA011013.1"		QY		987	TCCAAGCGGGCGGCTGGGTGTGGGTGACAGCTACGCCACCGTGGTGACACACGCCG 1046
/db_xref="gi:1754613"		Db		1398	TCCAAGCGGGCGGCTGGGTGTGGGTGACAGCTACGCCACCGTGGTGACACACGCCG 1457
/translation="MKESKNAATRKENGEPYELAKLLPLPSAITSOLDKASTIR LTTGYLKMRAVFPGLDAGWQSPRTGPLDSVAKELSHLLQTLIDGVFVVASDGRIM YISATSVLHGLSQVELTNGSIYIYIHPSDHDEMTAVLTAHPPLHLLHLOEYELERSF FLRMKCVLAKNAGLTCGKYVHCISGLIKIROYLMDLSYQIYGLVAVGQSLP PSATEIKLHNSNMEFRASLDLKLIFLDSRVTELTGYEPQDLIEKTLIHHVHGDFH LRYAHLHLLVGQVTKYRLSLKLGWVWQSVATVHNSRSRPHCIYVSNVYLTID VEYRELQSLDSQVSTKSQESWRVTLTSQETRSKAKPKNTKMTKLTNPYPPOYS SFQMDKLESCQVGNWRTSPPTNAVAPPEQQLHSEADLLYGPVSLPFSVYHGFPLD SHVFSKKRGLFPAKFGPOGSPCEVARFELTLPASSECQWHCANSIVPSSPAKNL SEPSVNAARHGLVNPYEPASAAARFCEPAPFSCQHYREAPALGPAKAPRQNS RDAARLALARPAPCEAPPAPEQAPQALPFLVLLNHRVLAARRPLGSAAPGAPEAAG SLRPRHPGPVAASAPGAPRPHYLGAASVITNGR"		QY		1047	TCGTCCCGGCGCCACCTGCATCGTGTAGTGTCAATATTATGTACTCACGAGATTGAATACAAG 1106
BASE COUNT		Db		1458	TCCTCCCGGCGCCACCTGCATCGTGTAGTGTCAATATTATGTACTCACGAGATTGTGNAFACAAG 1517
833 a 1110 c 952 g 717 t 2 others		QY		1107	GAACTTCAGCTGTCCTCGGAGCAGGTGTCCACTGCCAAGTCCCAGGACTCTCCAGGAGACC 1166
ORIGIN		Db		1518	GAACTTCAGCTGTCCTCGGAGCAGGTGTCCACTGCCAAGTCCCAGGACTCTCCAGGAGACC 1577
Query Match 46.3%; Score 1323.4; DB 10; Length 3614;		QY		1167	GCCTTGTCTACCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAG 1226
Best Local Similarity 88.2%; Pred. No. 7.4e-303;		Db		1578	ACCTTGTCTACCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAG 1637
Matches 1451; Conservative 0; Mismatches 191; Indels 3; Gaps 1;		QY		1227	ACAAGCTGAGAACAAACCTTACCCCCACAGCAATACAGCTCATTTCCAAATGGACAAA 1286
27 CCGGGCAGTGGGCTCCGCGGCTGGAGCAGCGCGGTCTAATATATGCCGCGGACCGA 86		Db		1638	ACAAGCTGAGAACAAACCTTACCCCCACAGCAATACAGCTCATTTCCAAATGGACAAA 1697
438 CAGCCCAAGCGGACTCCGCGGCTTGGAGCCAGCGAGGTCTATATGCCAGAGCGCA 497		QY		1287	CTGGAATCGCGCCAGCTCGGAAACTGGAGCCAGTCCCTGCAAGCGGTGCTGCTCCT 1346
87 GCGCGATGAAGAGAAGTCCAAGATCGGCCAAGACAGCAGGAGGAGAAGAAATGCG 146		Db		1698	CTGGAATCGCGCCAGCTCGGAAACTGGAGCCAGTCCCTGCAAGCGGTGCTGCTCCT 1757
498 GCGCGATGAAGAGAAGTCCAAGATCGGCCAAGACAGCAGGAGGAGAAGAAATGCG 557		QY		1347	CCAGAACTGCAGCCCACTCAGAAAGCAGTACCTTCTGTACAGCCATCCTACAGCCTG 1406
147 GAGTTTACAGCTTCCCAAGCTGCTCCGCTGCCGTCGCGCATCACTTCGAGCTGGAC 206		Db		1758	CCGGAAGTGCAGCCCACTCAGAAAGCAGTACCTTCTGTATGAGCCGCTACAGCCTC 1817
558 GAGTTTACAGCTGCGCAAGCTGCTCCGCTGCCGTCGCGCATCACTTCGAGCTGGAC 617		QY		1407	CCCTTCTCCTACCAATTTATGGACACTTCCCTTGGACTCTCACTTCTTCAGCAGCAAAAAG 1466
207 AAAGCGTCCATATCCGCTTCCACAGCAGTACCTGAAGATCGCGCCGCTCTTCCCGAA 266		Db		1818	CCCTTCTCCTACCAATTTATGGACACTTCCCTTGGACTCTCACTTCTTCAGCAGCAAAAAG 1877
618 AAAGCGTCCATATCCGCTTCCACAGCAGTACCTGAAGATCGCGCCGCTCTTCCCGAA 677		QY		1467	CCAATTTGCGCGCCAGTTCGGGCGAGCCCAAGGATCCCTTGTGAGGTGGCAGCCTT 1526
267 GGTTTAGGAGCGGTTGGGAGACAGCGAGCGCGCGCGCTGACGCGCTGCGCAAG 326		Db		1878	CCAATTTGCGCGCCAGTTCGGGCGAGCCCAAGGATCCCTTGTGAGGTGGCAGCCTT 1937
678 GGTCTAGGAGCGGTTGGGAGACAGCGAGCAGGAGCGGCTCTGGACAGCTGGCAAA 737		QY		1527	TTCTCTGAGCACAATGCCAGCCAGCGGTGAATGCCAGTGGCAATATGCCAACCCCTTAGTG 1586
327 GAGCTGGATCGCACTTGTGTCAGACTTTGGATGGATTTGTTTTGTGTAGCATCTGAT 386		Db		1938	TTCTCTGAGCACAATGCCAGCCAGCGGTGAATGCCAGTGGCAATATGCCAACCCCTTAGTG 1997
738 GAGCTGGATCGCACTTGTGTCAGACTTTGGATGGATTTGTTTTGTGTAGCATCTGAT 797		QY		1587	CCTAGCAGCTGCTCCAGCTAAAATCTCTCAGAGC---CACCGCGCAACCTCTCTAGG 1643
387 GGCAAAATCATATATATCCGAGACCGCTTGTCTCATTTAGCTTATCCAGGTGGAG 446		Db		1998	CCAGCAGCTCATCGCAGCTAAAACCTTTCTGAGCCTTCTCTGTGATGCTGCCCGG 2057
798 GGCAAAATCATATATATCCGAGACCGCTTGTCTCATTTAGCTTATCCAGGTGGAG 857		QY		1644	CACAGCTGTGCCAGCTACGAAG 1668
447 CTCAGCGCAACAGTATTATGATATACATCCTTCTGACCAACAGATGAGTACCGCT 506		Db			
858 CTCAGCGCAACAGTATTATGATATACATCCTTCTGACCAACAGATGAGTACCGCT 917		QY			
507 GTCTCTCAGCGCCACCGCGCTGCACACACCTGCTCCAAAGATGATAGAGAGG 566		Db			
918 GTACTTACAGCCACCGCTCCACCATCACCTGCTCCAAAGAGTACGAGATAGAGCGG 977		QY			





	BASE COUNT	700 a	968 c	804 g	599 t	
	ORIGIN					
	Query Match	46.2%	Score 1320.2:	DB 10;	Length 3071;	
	Best Local Similarity	88.1%;	Pred. No. 4.2e-302;			
	Matches 1449;	Conservative	0;	Mismatches 193;	Indels 3;	Gaps 1;
QY	27	CGGGGCGAGTGGGGCTCCGGGGCCCTGGAGCAGCGCGGGTCTAAATATGCCGAGCGCGA	86			
Db	163	CAGCCCAAGCGGGGACATCCGGGGCCCTGGAGCGCAGCGAGGTCTTAATATGCCAGAGCGCA	222			
QY	87	GGCGCGATGAAGGAGAAGTCCAAGATTCGGCGCCAAAGCACAGGAGGGAGAGGAAATGGC	146			
Db	223	GGCCCGATGAAGGAGAAGTCCAAATAATCGCGGCCAAACACAGGAGGGAGAGGAAATGGC	282			
QY	147	GAGCTTTACGAGCTTGCCAAAGCTGCTCCCGCTGCGCTCGGCCATCACTTCGCAGCTGGAC	206			
Db	283	GAGTTTTACGAGCTGGCCAGCTGCTCCCGCTGCCTTCGGCCATCACTCGCAGCTGGAC	342			
QY	207	AAAGCGTCCATATATCCGCCCTACACAGAGCTACTGAAGATGCGCGCGCTCTTCCCGCGAA	266			
Db	343	AAAGCGTCCATATATCCGAGCTACACACAGAGTACTGAAGATGCGCGCGCTCTTCCCGCGAA	402			
QY	267	GCTTTAGGAGACGGCTGGGAGACAGCCGAGCGCGCGGGCCCTTGGAGCGCGCTGCCAAG	326			

[illegible][illegible]





Db	1303	GTACTTACAGCCACCCAGCGCTCCACCATCATCGCTGCTCCAAAGAGTACGAGATAGAGCGG	1362
Qy	567	TCGTTCTTTTTCGAATGAATGTCTTGGCGAAAGACGCGGCGCTGACCTGCGAGC	626
Db	1363	TCCTTTCTTTTCGAATGAATGTCTTGGCGAAAGACGCGGCGCTGACATGACGT	1422
Qy	627	GGATACAGGTTCATCCACTGCGAGTGGCTACTTGAAGATCAGGCAGTATATGCTGGACATG	686
Db	1423	GGATACAGGTTCATCCACTGCGAGTGGCTACTTGAAGATCAGGCAGTATATGCTGGACATG	1482
Qy	687	TCCCTGTACGACTCCTCTGCTACCATGTTGGGGGTGGTGGCCGTGGGCGAGTCTGCTGCCA	746
Db	1483	TCCCTGTACGACTCCTCTGCTACCATGTTGGGGGTGGTGGCCGTGGGCGAGTCTGCTGCCA	1542
Qy	747	CCAGTGCCTATCCAGGATCAAGCTGTACAGTAACTGTTTCATGTTTCAGGCGCCAGCTT	806
Db	1543	CCAGTGCCTATCCAGGATCAAGCTGTACAGTAACTGTTTCATGTTTCAGGCGCCAGCTT	1602
Qy	807	GACCTGAAGCTGATATTCCTGATTCAGGGTGCAGGAGTGCAGGTTACGAGCCGAG	866
Db	1603	GACCTGAAGCTGATATTCCTGATTCAGGGTGCAGGAGTGCAGGTTACGAGCCGAG	1662
Qy	867	GACCTGATCGAGAGACCTATACATCAGCTGCAGGCTGCGAGTGTTCACCTCGGC	926
Db	1663	GACCTGATCGAGAGACCTATACATCAGCTGCAGGCTGCGAGTGTTCACCTCGGC	1722
Qy	927	TACGCACACACCTCTCTGTTGTAAGGCGCAGCTACCCACCAAGTACTACCGGCTGCTG	986
Db	1723	TATGCACACACCTCTCTGTTGTAAGGCGCAGCTACCCACCAAGTACTACCGGCTGCTG	1782
Qy	987	TCCAGCGGGGCGGCTGGTGTGGTGCAGAGTACGCCACCGTGGTGCACACAGCGCG	1046
Db	1783	TCCAGCTGGGGGCTGGTGTGGTGCAGAGTACGCCACCGTGGTGCACACAGCGCG	1842
Qy	1047	TGTCCTCCGCGCCACTGATCGTGTAGTGTCAATATGTTACTCAGGAGATTAATACAA	1106
Db	1843	TCCTCCGCGCCACTGATCGTGTAGTGTCAATATGTTACTCAGGAGATTAATACAA	1902
Qy	1107	GAACCTTACGCTGCTCCTGGAGAGCTGTCCACTGCCAGTCCAGAGCTCTCTGGAGGACC	1166
Db	1903	GAACCTTACGCTGCTCCTGGAGAGCTGTCCACTGCCAGTCCAGAGCTCTCTGGAGGACC	1962
Qy	1167	GCTTTGTCTACTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	1226
Db	1963	ACCTTTGTCTACTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	2022
Qy	1227	ACAAAGCTG-AGAAACAAACCTTACCCCGCAGCAATACAGCTCATTCCTCAATGGACAA	1285
Db	2023	ACAAAGCTGAGAAACAAACCTTACCCCGCAGCAATACAGCTCATTCCTCAATGGACAA	2074
Qy	1286	ACTGGAATGCGCCAGCTCGGAAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCC	1345
Db	2075	ACTGAGTGCACCGCAGGTGGAACTGGAGAACTAGTCCCTCCCAAAATGCTGCTGCTCC	2134
Qy	1346	TCCAGAACTGACGCCCCCTCAGAAAGAGTGAACCTTGTGTACAGCCATCTCAGAGCT	1405
Db	2135	CCCGGAACAGCAGCTCCATTCAGAAAGCCAGCGACCT--TATATGGGCGCCCTACAGGCT	2192
Qy	1406	GCCCTTCTCTACCATATGAGACCTTCCTCTGGACTCTCACTTCTTCAGCAGCAAAA	1465
Db	2193	CCCTTCTCTCTACCATATGAGACCTTCCTCTGGACTCTCACTTCTTCAGCAGCAAAA	2252
Qy	1466	GCAATATGTTGCGGCGCAAGTTCGGGCGAGCCCAAGATCCCTTGTGAGGTGGCAGCTT	1525
Db	2253	GCCAGATTTGCCGCGCAAGTTCGGGCGAGCCCAAGATCCCTTGTGAGGTGGCAGCTT	2312
Qy	1526	TTTCTGTGACCAATGCGCAGCAGCGGTGAATGCGCAGTGGCAATTATGCCAACCCCTAGT	1585
Db	2313	CTTCTGTGACCAATGCGCAGCAGCGGTGAATGCGCAGTGGCAATTATGCCAACCCCTAGT	2372
Qy	1586	GCCTAGCAGCTCGTCTCCAGCTTAAATTCCTCCAGAGC---CACCGGCGCAACACTGCTAG	1642
Db	2373	GCCAGCAGCTCATGCGCAGCTTAAACACCTTCTGAGCCCTTCTGAGCTGCTGCTGCTG	2432
Qy	1643	GCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1668
Db	2433	GCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2458
RESULT 13	AB063285/c	250000 bp	linear
LOCUS	AB063285	250000 bp	DNA
DEFINITION	Homo sapiens HLCS gene for holocarboxylase synthetase, complete cds.		
ACCESSION	AB063285		
VERSION	AB063285.1	GI:15823776	
KEYWORDS			
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Yang, X., Aoki, Y., Li, X., Sakamoto, O., Hiratsuka, M., Kure, S., Taheri, S., Christensen, E., Inui, K., Kubota, M., Ohira, M., Ohki, M., Kudoh, J., Kawasaki, K., Shibuya, K., Shintani, A., Asakawa, S., Minoshima, S., Shimizu, N., Narisawa, K., Matsubara, Y. and Suzuki, Y.		
TITLE	Structure of Human Holocarboxylase Synthetase Gene and Mutation Spectrum of Holocarboxylase Synthetase Deficiency		
JOURNAL	Hum. Genet. (2001) In press		
REFERENCE	2 (bases 1 to 250000)		
AUTHORS	Shimizu, N., Kudoh, J. and Shibuya, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JUN-2001) Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-5363-3754, Fax: 81-3-3351-2370)		
COMMENT	This sequence is comprised of as follows, 1-46502 bp; KB318C2 (Acc#AP000704), 46503-88043 bp; KUD37D12 (Acc#AP000703), 88044-89026 bp; Q63D8 (Acc#AP000702), 89027-128111 bp; Q41G5 (Acc#AP000701), 128112-164047 bp; Q2F5 (Acc#AP000700), 164048-199476 bp; KUD66G11 (Acc#AP000699), 199477-232385 bp; KUD8F6 (Acc#AP000698), 232386-250000 bp; KB594G10 (Acc#AP000697).		
FEATURES	Location/Qualifiers		
source	1...250000		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="21q22.1"		
	/note="assembled with 8 independent clone"		
gene	4950..243990		
	/gene="HLCS"		
mrna	join(4950..5028,33016..33150,49032..49177,56192..56354,57794..58737,64753..64935,98052..98323,227897..227964,230010..230170,235341..235455,238427..238640,240765..241251)		
	/gene="HLCS"		
	/note="mRNA sequence is reported in Acc#D23672"		
exon	4950..5028		
	/gene="HLCS"		
	/number=1		
mrna	join(28713..28742,33016..33150,46899..47853,56192..56354,57794..58737,64753..64935,98052..98323,227897..227964,230010..230170,235341..235455,238427..238640,240765..243990)		
	/gene="HLCS"		
	/note="mRNA sequence is reported in Acc#D87328"		
exon	28713..28742		
	/gene="HLCS"		
	/number=2		
exon	32762..33150		
	/gene="HLCS"		
	/number=3		
exon	33016..33150		
	/gene="HLCS"		
	/number=3		

mRNA join(33077..33150,49032..49177,56192..56354,57794..58737,64753..64935,98052..98323,227897..227964,230010..230170,235341..235455,238427..238640,240765..240978)  
/gene="HLCS"  
/note="mRNA sequence is reported in Acc#X80160"  
46899..47853  
/gene="HLCS"  
/number=4  
49032..49177  
/gene="HLCS"  
/number=5  
56192..56354  
/gene="HLCS"  
/number=6  
227897..227964,230010..230170,235341..235455,238427..238640,240765..240978)  
/gene="HLCS"  
/codon\_start=1  
/product="holocarboxylase synthetase"  
/protein\_id="BAB68550.1"  
/db\_xref="GI:1582377"  
/translation="MEDRLHMDNCLVPOKIVSVHLQDSTLKEVDQVSNKQAOILEPK  
PEPSLEIKPQDQMEHVGRDPRALGEEPKQRRGSASGSEPADSDRGSGPVEHYHLH  
LSSCHECELENSTIESVKFAENIPDLIDYSSLESVADETSPEREGRRVNLTK  
APNLLTVGSDSQEALGRFHEVRSVLADCDIDSYILYHLLDSALRDPMDNCLLV  
IATRESIPDLQKFMAYLSQGGKVLGSSFFGFOVTSKGLHKTQNLVFSKAD  
QSEVLSVSGCRYOEGPVLSPGRLOHLENDKDMIVHPVFTGRGEAVLCQVH  
LELPSSNIQVOTPEDENLLKSNFRBYEVLEILTLGLSCDMKQWPAITPLYLSSAA  
EETRDPLMOLGKHVSEGEIKSQSLRFVSIVSEVEITPSCPVPVWNMEAFSEH  
FNLVIRQNLQKGLVILFAEVTPTTMRLLDGLMFOQPMGLIVIAARQTEGGR  
GNWVSPVQALUSTLISIPRLSQLRPIFYOHLSMVAVVEAVRSIPEYODINLVR  
KPNDIYSDLMKIGVNLSTLMGETFYILIGCFNVTNSNPTICINDLITEYNKQH  
KAEKPLRADYLARVVTVLEKLIKERQDKGPNVPLXYRYRVVHSGQVHLGSAREP  
KVSIVGLDSDGFLQVHQEGEVTVTHPDGNSFDMLRLNLILPKRR"  
57794..58737  
/gene="HLCS"  
/number=7  
64753..64935  
/gene="HLCS"  
/number=8  
98052..98323  
/gene="HLCS"  
/number=9  
227897..227964  
/gene="HLCS"  
/number=10  
230010..230170  
/gene="HLCS"  
/number=11  
235341..235455  
/gene="HLCS"  
/number=12  
238427..238640  
/gene="HLCS"  
/number=13  
240765..243990  
/gene="HLCS"  
/number=14  
complement(247090..247517)  
/gene="SIM2"  
complement(247090..247517)  
/gene="SIM2"  
/number=11  
BASE COUNT 64877 a 54574 g 56864 g 73685 t  
ORIGIN

Query Match 40.0%; Score 1142.8; DB 9; Length 250000;  
Best Local Similarity 99.8%; Pred. No. 6.7e-260;  
Matches 1144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1714 CCGGAGCCGAGCAGCCATGGCGTGGTGGCAGATGGAGACAGAACCCCTCAGCCTTG 1773  
|||||

Db 250000 CCGGAGCCGAGCAGCCATGGCGTGGTGGCAGATGGAGACAGAACCCCTCAGCCTTG 249941  
QY 1774 GGCAAACTTCCCTCTTTCTTCTTAAGTAGGGCTTCTGTGCTTCTTCTCAATG 1833  
Db 249940 GGCAAACTTCCCTCTTTCTTCTTAAGTAGGGCTTCTGTGCTTCTTCTCAATG 249881  
QY 1834 CAGGTGCTCTCGAGAGTGAGAAATGGCAGTCTGCTGCTCGGGGACACTAGTGACAGT 1893  
Db 249880 CAGGTGCTCTCGAGAGTGAGAAATGGCAGTCTGCTGCTCGGGGACACTAGTGACAGT 249821  
QY 1894 ATAAAGGGCAAGGAAACCGAGTATCTGGCCTTTCAGTAAATCTCTGGGCCACATTCACCA 1953  
Db 249820 ATAAAGGGCAAGGAAACCGAGTATCTGGCCTTTCAGTAAATCTCTGGGCCACATTCACCA 249761  
QY 1954 ACCAAAGGGGACAGTGATTTTCAAAACAGAGTCCCATGCTGTGAGAACACCCAGCTGC 2013  
Db 249760 ACCAAAGGGGACAGTGATTTTCAAAACAGAGTCCCATGCTGTGAGAACACCCAGCTGC 249701  
QY 2014 ATTTCTTTTGGCAAGATTCCTTTCCACTCCAACAGAGTGAATATTTGAGACAAACGGCC 2073  
Db 249700 ATTTCTTTTGGCAAGATTCCTTTCCACTCCAACAGAGTGAATATTTGAGACAAACGGCC 249641  
QY 2074 TATTGGCTATTTCCCATGCCAGTTTGGAAAGTGGGAAACATATGGTGAAATTTTGGG 2133  
Db 249640 TATTGGCTATTTCCCATGCCAGTTTGGAAAGTGGGAAACATATGGTGAAATTTTGGG 249581  
QY 2134 GCTTGGGGACAGAAATGCCACTCACCAAGGCAAGAAACACAAACCCCTCCAGGCT 2193  
Db 249580 GCTTGGGGACAGAAATGCCACTCACCAAGGCAAGAAACACAAACCCCTCCAGGCT 249521  
QY 2194 CAGTTTCTTCACTGTAAATGGGTGAAGTGTGATGTCCTACTCCAGAGCACAC 2253  
Db 249520 CAGTTTCTTCACTGTAAATGGGTGAAGTGTGATGTCCTACTCCAGAGCACAC 249461  
QY 2254 ACACAGTAGGGACCTGCCCTGTACATGTAGTTCAACAGAAAGAAATGGCTTTTCCACTT 2313  
Db 249460 ACACAGTAGGGACCTGCCCTGTACATGTAGTTCAACAGAAAGAAATGGCTTTTCCACTT 249401  
QY 2314 CTCCTTGGTGGCAGGCAAGCAGATGCTCTCGGGAGATACCGCCAGCTCCCGAGACGCA 2373  
Db 249400 CTCCTTGGTGGCAGGCAAGCAGATGCTCTCGGGAGATACCGCCAGCTCCCGAGACGCA 249341  
QY 2374 GACTGACTCTCTTGTCTGCTGGACCAACCCAGCAGAGAGGTGGAAGTGGGACACAGA 2433  
Db 249340 GACTGACTCTCTTGTCTGCTGGACCAACCCAGCAGAGAGGTGGAAGTGGGACACAGA 249281  
QY 2434 GGTTAGCTGCAGGACATGTATTCCTCATTCACCGAGACCTAACTGCCGCTCAGAGTGA 2493  
Db 249280 GGTTAGCTGCAGGACATGTATTCCTCATTCACCGAGACCTAACTGCCGCTCAGAGTGA 249221  
QY 2494 GACCGAGATGGTGCAGATGCCCTGCAGTCCCATTTAAATGTGGGTGAAGGTGACATCAGGA 2553  
Db 249220 GACCGAGATGGTGCAGATGCCCTGCAGTCCCATTTAAATGTGGGTGAAGGTGACATCAGGA 249161  
QY 2554 TTATGTGCCCGCAGGCGGGCTCAGTGGCTCACACTGTAAATCCCGACACTTTTGGAGGCC 2613  
Db 249160 TTATGTGCCCGCAGGCGGGCTCAGTGGCTCACACTGTAAATCCCGACACTTTTGGAGGCC 249101  
QY 2614 AAGTGGCGGATCACCTGAGGTGAGGTTGGCAAGCCTGCCAACAGCTGAAACC 2673  
Db 249100 AAGTGGCGGATCACCTGAGGTGAGGTTGGCAAGCCTGCCAACAGCTGAAACC 249041  
QY 2674 CCATCTCCACTTAAATAACAAAAATAGTTGGGCATGGTGGGACACCTGTATATCCAG 2733  
Db 249040 CCATCTCCACTTAAATAACAAAAATAGTTGGGCATGGTGGGACACCTGTATATCCAG 248981  
QY 2734 CTACTCTGGAGGTGAGATAGGAGATCACTTTGAACCCGGAGGTGAGGTGCAGTGAG 2793  
Db 248980 CTACTCTGGAGGTGAGATAGGAGATCACTTTGAACCCGGAGGTGAGGTGCAGTGAG 248921  
QY 2794 CTAGATCAGATCAGTCCAGCTCCAGCTGGGTGAACAGAGTGAGACTGTCTCAAAAAAAA 2853  
Db 248920 CTAAGATCAGATCAGTCCAGCTCCAGCTGGGTGAACAGAGTGAGACTGTCTCAAAAAAAA 248861





Mon Jul 21 15:46:54 2003

us-09-923-684-2.rge

983 AACCTCAGGACCTGATTGAGAGACCTCTGTACACCATGTGCACGGCTCGGACACCTCC 1042  
919 ACCTCCGCTACGACACACCTCTCTGTGTGAAGGCCAGGTACACCAAGTACTACC 978  
1043 ACCTGGCGCTGGGGACCATTTGCTGCTGTGAAGGACAGTGCACCAAGTACTACA 1102  
979 GGCTGCTGTCCAAAGCGGGGGTGGTGTGGTGCAGAGCTACGCCACCGTGGTGACA 1038  
1103 GGTTCCTGGGGAACACGGCGGTGGTATGGTGAAGCTACGGGACCATCGTGACA 1162  
1039 ACAGCCGCTGTCCCGGCCCTACCTGATCGTGTGATGCTCAATTATGTACTCACGGAGATTG 1098  
1163 ACAGTCGCTCTCCAGCCACATGTATCGTCAGCGTCACTATGTCTCTCCACAGACACAG 1222  
1099 AATACAAAGGAACTTCAGCTGTCCCTGGAGCAGGTGTCCACTGCCAAAGTCCAGGACTCCT 1158  
1223 AATACAAAGGCTGCAGCTCTCCCTGGATCAGATCTCAGGCTCCAA---ACCAGCCTTCT 1279  
1159 GGAGGACCGCTTGTCTACCTACAGAAACTAGGAATTTAGTGAACCCCAAAATACCA 1218  
1280 CCTATACGACGAGCTCCACCCCACTGATGATGACAAAGGGGGCCCAAAATCCCGGC 1339  
1219 AGATGAAGACAAAGCTGAGAAACACCTTACCCCCACACACATACAGCTCATTTCCAAA 1278  
1340 TCTCCAGCTCAAGTCAAAATCCAGGACTTCCCATACCTCAGTATTCGGGATTTACACA 1399  
1279 TGGACAAACTGGAATGGCGGCGAGCTCGGAAACTGGAGAGCCAGTCCCTTGCAGCGCTG 1338  
1400 CAGAAAGATCGGAATCTGATCATGACAGCCAGTGGGGCGGAAGTCCCTTGACCGACACGG 1459  
1339 CTGCTCTCCAGAACTGACGCCACCTCAGAAAGCAGTGCCTTCTGTACACGCCATCCT 1398  
1460 COTCTCCGAGCTTGTGACCCCGCGGATAGGCTGGCTCCACACAGCCGATCSTGG 1519  
1399 ACAGCTGCGCTTCTCTTACCATTTATGGACACTTCCCTCTGGACTCTCAGCTTTTACGA 1458  
1520 CTTACAGACAGTTTTCGGACCGCAGCTCTCTCTGTATGGCTTTGGCGTTGACCACTCGA 1579  
1459 GCAAAAGCCAATGTTCCGGCCAGTTTCGGGACCCCAAGGATCCCTTGTGAGGTGG 1518  
1580 GGCTGGTGAAGAGAGGCAATTTCCATACCCAGGCTGTGAAGGAGCGCGATGTGAGGCAG 1639  
1519 CACGCTTTTCTCTGAGCAAAATGCCAGCAGCGGTGAATGCGCAGTGGCAATTTGCCAAC 1578  
1640 GCAGGTACTTCTGGAGCGCGCGAGCGCGGAGGAGCCCTGTGGGGCTCTCGCGCAG 1699  
1579 CCCTAGTGCCTAGCAGCTCGTCTCCAG 1605  
1700 CCTTGGCCCTGACAAAGGCTCCCGAG 1726

Search completed: July 17, 2003, 19:54:13  
Job time : 7282 secs